

Franco Zavatti

Appunti di
ESPERIMENTAZIONI DI FISICA I

Corso di Laurea in Astronomia

a.a. 2006-2007

INDICE

1. INTRODUZIONE	5
2. PROBABILITÀ	6
<i>Definizioni di Probabilità</i>	8
<i>Somma e prodotto di eventi</i>	9
<i>Teorema di moltiplicazione</i>	11
<i>Probabilità totale</i>	13
<i>Teorema delle ipotesi (Bayes)</i>	14
<i>Teoremi delle prove ripetute</i>	15
3. LEGGI DI DISTRIBUZIONE	17
<i>Funzione di distribuzione</i>	18
<i>Probabilità di appartenenza ad un intervallo</i>	19
<i>Densità di probabilità</i>	20
<i>Caratteristiche numeriche</i>	22
<i>Proprietà del valore di aspettazione</i>	25
<i>Proprietà della varianza</i>	25
<i>Funzione generatrice dei momenti</i>	27
<i>Teorema dell'asse parallelo</i>	28
<i>Disuguaglianza di Chebychev</i>	29
4. DISTRIBUZIONI DI PIÙ VARIABILI	32
<i>Due variabili</i>	32
<i>Momenti</i>	34
<i>Funzioni di N variabili</i>	39
<i>Momenti</i>	40
5. TRASFORMAZIONI DI VARIABILI	42
<i>Trasf. lineari - Propagazione errori</i>	46
6. FUNZIONI DI DISTRIBUZIONE PIÙ COMUNI	48
<i>Bernoulli</i>	48
<i>Binomiale</i>	49
<i>Multinomiale</i>	53
<i>Legge dei grandi numeri</i>	54
<i>Poisson</i>	55
<i>Uniforme</i>	59
<i>Gauss (normale)</i>	61
χ^2	66
<i>Student</i>	67
<i>Fisher</i>	68
<i>Teorema Limite Centrale</i>	69
7. CAMPIONAMENTO	71
<i>Stima</i>	72
<i>Stima dei parametri della regr. lineare</i>	76
<i>Campionamento di una gaussiana - χ^2</i>	77
8. MASSIMA VEROSIMIGLIANZA	78
<i>Fit con la massima verosimiglianza</i>	81
<i>Stime per intervallo</i>	83
9. TEST DI IPOTESI STATISTICHE	86
<i>test di normalità per medie</i>	87

	<i>t</i> - test per medie (Student)	87
	<i>F</i> - test per varianze (Fisher)	91
	χ^2 - test per adattamento curve	93
10.	TAVOLE
	<i>Curva normale standard</i>	96
	<i>Quantili della curva normale</i>	97
	χ^2	98
	<i>Student</i>	99
	<i>Fisher</i>	100

Scopo del corso

Scopo del corso è fornire gli elementi di base del calcolo delle probabilità e della statistica, applicate all'analisi dei dati sperimentali. Si vuole anche fornire una conoscenza relativa alla preparazione ed attuazione di esperienze di laboratorio con relativa analisi statistica dei risultati. *Statistica e Analisi dei dati sperimentali*

Probabilità: Somma e prodotto di eventi, teorema di moltiplicazione, Probabilità totale, teorema di Bayes, teorema delle prove ripetute. *Leggi di distribuzione:* funzione di distribuzione, probabilità di appartenenza ad un intervallo, densità di probabilità, caratteristiche numeriche, proprietà del valore di aspettazione, proprietà della varianza, funzione generatrice dei momenti, teorema dell'asse parallelo, disuguaglianza di Chebyshev. *Distribuzioni di più variabili aleatorie:* funzioni di due variabili e loro momenti, funzioni di N variabili e loro momenti. *Trasformazioni di variabili:* trasformazioni lineari, propagazione degli errori. *Funzioni di distribuzione più comuni:* bernoulli, binomiale, multinomiale, legge dei grandi numeri, poisson, uniforme, distribuzione normale, χ^2 , Student, Fisher, teorema limite centrale. *Campionamento:* Stima, criteri per una buona stima, stima dei parametri dei minimi quadrati, campionamento di una distribuzione gaussiana - χ^2 . *Massima verosimiglianza:* generalità, interpolazione con la massima verosimiglianza, media pesata. *Test di ipotesi statistiche:* generalità, test di normalità per medie, t-test per medie (Student), F-test per varianze (Fisher), χ^2 -test per confronto tra dati sperimentali e leggi teoriche.

Misure di grandezze meccaniche: misure di lunghezza, di tempo, di masse

Misure di radioattività: misure di efficienza, di assorbimento

Misure ottiche: cenni di ottica geometrica, misure di distanze focali di lenti convergenti e divergenti, misure su sistemi ottici;

La frequenza alle 5 prove di laboratorio che fanno parte integrante del corso è obbligatoria

Testi indicati

FRANCO ZAVATTI, *Appunti di Esperimentazioni di Fisica I*, 2003. Dispensa da integrare con Brandt e Ventsel.

FRANCO ZAVATTI, *Prove di Laboratorio*, 2003. Dispensa.

SIEGMUND BRANDT, *Statistical and Computational Methods in Data Analysis*, 1970, North Holland Publishing Company. Capitoli 3-8

ELENA VENTSEL, *Teoria delle Probabilità*, 1983, Edizioni MIR. prime 80 pagine

Testi consigliati per approfondimenti

G. VICARIO, R. LEVI, *Calcolo delle probabilità e statistica per ingegneri*, 1997, Bologna, Società Editrice Esculapio.

M. BRAMANTI *Calcolo delle probabilità e statistica*, 1997-2001, Bologna, Società Editrice Esculapio.

G. CICCHITELLI, *Probabilità e Statistica*, 1994, Rimini, Maggioli Editore

G. CICCHITELLI, M. A. PANNONE, *Complementi ed esercizi di Statistica descrittiva ed inferenziale*, 1994, Rimini, Maggioli Editore

C. CAMETTI, A. DI BIASIO, *Introduzione all'elaborazione dei dati sperimentali*, 1994, Roma, CISU. Molto simile a Brandt.

Testi consigliati per approfondimenti sugli strumenti di laboratorio

GIOVANNI VALDRÈ, UGO VALDRÈ *Misure e complementi di Fisica*, Bologna, CLUEB, 1998.

ETTORE PANCINI, *Misure ed apparecchi di Fisica*, Roma, Libreria Eredi Virgilio Vaschi.

G.CORTINI, S. SCIUTI, *Misure ed apparecchi di Fisica*, Roma, Libreria Eredi Virgilio Vaschi.

**Per me niente è sicuro quanto la cosa incerta;
niente è oscuro se non ciò che è evidentissimo;
non dubito, se non di ciò che è certo;
la conoscenza mi deriva da fatti casuali.**

**Rien ne m'est sûr que la chose incertaine ;
Obscur, fors ce qui est tout évident ;
Doute ne fais, fors en chose certaine ;
Science tiens à soudain accident ;**

Francois Villon: Ballade du concours de Blois

1. INTRODUZIONE

Scopo del corso è lo studio del modo di **trattare i dati sperimentali per ottenere quello che noi considereremo il valore più probabile del risultato delle misure**. Quando ci si accinge a studiare un fenomeno di qualsiasi natura (*fisica, chimica, sociale, letteraria*) la prima cosa da fare è raccogliere informazioni, cioè **fare misure** ed **interpretare** i dati raccolti. Interpretare significa quasi sempre :

- definire, da quanto osservato, un comportamento (più) generale che possa essere seguito da altre serie di dati, presi in condizioni e in tempi differenti, dello stesso fenomeno o, viceversa,
- verificare che i dati attuali si inquadrino nel comportamento generale ottenuto da studi precedenti.

Fare misure significa ottenere un risultato (numerico o non) in corrispondenza di una grandezza che si fissa a priori.

Es.: ottenere la posizione di un oggetto mobile a tempi fissati; osservare come cambia la luce tra due inquadrature di un film; per ogni persona intervistata in un sondaggio, registrare le risposte, ecc.

Misurare è difficile, e lo è soprattutto mantenere le condizioni esterne invariate (Es. quando gli studenti entrano in laboratorio la temperatura ambiente cresce ed alcuni tipi di misure possono essere falsate).

In generale, le misure di grandezze fisiche ripetute, anche le più semplici, quasi mai danno lo stesso risultato numerico: ciò è dovuto al fatto che le operazioni che si compiono per eseguire una misura sono più o meno complicate ed il modo stesso di operare introduce delle variazioni nel risultato.

Distinguiamo quindi subito due cause che influenzano il risultato di una misura:

- a) *Deviazioni sistematiche*, che rendono il risultato errato in modo tale che ripetendo la misura si sbaglia sempre allo stesso modo. Cause di questo tipo sono: strumenti non calibrati, condizioni di lavoro diverse da quelle per cui lo strumento adoperato è stato progettato o è supposto essere adoperato. Di queste *cause* (e perciò di questi errori) non si può fare una teoria, ma piuttosto un'elencazione che deriva quasi esclusivamente dall'esperienza, e comunque sono tali che, una volta individuate, possono essere rimosse o si può correggere il risultato. Non ne parlerò estesamente, ma vorrei soltanto ricordare che, poichè a volte possono falsare di molto il risultato di una misura, tutte le volte che si fa una misura bisogna, nei limiti del possibile, essere sicuri di non introdurre errori sistematici.
- b) *Deviazioni casuali*, che sono tali da non rendere il risultato di una misura ripetibile, ma normalmente sono sufficientemente piccole e, **cambiando di volta in volta**, non sono prevedibili ma possono essere studiate e

trattate dal punto di vista statistico, qualora si disponga di un certo numero di misure.

Spenderemo una parte del corso per imparare a trattare i risultati delle misure che si presentano all'osservatore come se fossero **variabili aleatorie**, intendendo con ciò una variabile il cui successivo valore osservato non è predeterminato ma è tale che se si esegue un grande numero di osservazioni, la probabilità che il valore della variabile sia in un certo intervallo è determinata da una legge chiamata densità di probabilità che indicheremo con $f(x)$.

Ci sono poi altri problemi caratteristici dell'analisi dei dati sperimentali che dovremo affrontare.

- a) **Test delle ipotesi** che consiste nello stabilire se un certo risultato sperimentale è in accordo con una certa teoria od ipotesi.
- b) **Discriminazione fra diverse cause** - Si studia l'influenza su di un campione statistico di varie cause applicate per osservare se i risultati sono diversi o no.
- c) **Si stabiliscono** i parametri di leggi fisiche conosciute a partire dai risultati delle misure.
- d) **Si cerca di stabilire quale espressione analitica** abbiano certi legami fra grandezze, con lo scopo di definire quale legge fisica li governi.

Partiremo quindi con lo studio delle funzioni che descrivono il comportamento delle variabili aleatorie, cioè **le densità di probabilità**.

Poi vedremo come da un **campione statistico** si possano ricavare **le stime dei parametri** che caratterizzano le densità di probabilità.

Passeremo successivamente ad analizzare **i test delle ipotesi** e applicazioni varie.

2. PROBABILITÀ

Tutte le variazioni che osserviamo sono dovute ad effetti secondari che influenzano i risultati di un esperimento, ma che non possono essere considerati essenziali.

Le condizioni essenziali restano, grosso modo, costanti e definiscono l'apparire di un esperimento. Le condizioni secondarie variano da un esperimento all'altro e generano le variazioni aleatorie nei risultati.

Per studiare sempre meglio un fenomeno reale, si tenta di produrre dapprima una legge generale che, definendo un a situazione esterna relativamente semplice, possa prevedere il verificarsi del fenomeno. Poi il dettaglio diviene più fine e si prendono in considerazione effetti "secondari" via via meno importanti. Nella realtà dei fatti, però, gli effetti secondari sono talmente numerosi che la combinazione della loro azione può avere peso non trascurabile nell'andamento della prova e nello stesso tempo essere troppo complessa per essere studiata in modo deterministico. Gli **elementi casuali** indeterminati che contraddistinguono queste misure **richiedono metodi di studio diversi da quelli**

deterministici. Questi metodi formano il corpo della Teoria della Probabilità, che studia le regolarità degli eventi aleatori.

In questo studio si è aiutati dal fatto sperimentale che all'aumentare degli esperimenti, di qualunque natura essi siano, la maggior parte dei risultati si addensa attorno ad un'area ristretta, diminuendo con la distanza da questa area e disponendosi in modo simmetrico rispetto al punto di massimo addensamento.

In pratica succede che un grandissimo numero di eventi aleatori si combina per dare un risultato che non è più aleatorio.

La scala delle probabilità si estende tra l' evento certo con probabilità uguale ad 1 e l' evento impossibile con probabilità uguale a 0.

Gli eventi possono formare un GRUPPO COMPLETO *se almeno un evento deve accadere* Es. testa o croce lanciando una moneta.

Oppure possono essere:

INCOMPATIBILI *se nessuna coppia di eventi può verificarsi contemporaneamente.* Es. testa e croce lanciando una moneta.

EQUIPROBABILI *se hanno la stessa probabilità di apparire.* Es. apparizione di un numero da 1 a 6 lanciando un dado. La probabilità dell'evento A è data da $P(A) = m/n = \text{favorevole}/\text{totale}$.

ESEMPI:

1) In un'urna ci sono 2 palline bianche e 3 nere. Si estrae 1 pallina. Qual'è la probabilità che sia bianca?

Soluzione:

$$P = m/n = 2/5$$

2) In un'urna ci sono a palline bianche e b nere. Si estraggono 2 palline. Qual'è la probabilità che siano bianche ?

Soluzione:

B= evento = estrazione di 2 palline bianche

$$n = C_{a+b}^2 = \frac{(a+b)!}{2!(a+b-2)!}$$

$$m = C_a^2 = \frac{a!}{2!(a-2)!}$$

$$P(B) = \frac{C_a^2}{C_{a+b}^2} = \frac{2!(a+b-2)!}{(a+b)!} \frac{a!}{2!(a-2)!}$$

3) In un portacenere vengono deposti fiammiferi usati

a) da 10 fumatori di sigarette

b) da 4 fumatori di pipa

Se è noto che i fumatori di pipa consumano 5 volte più fiammiferi di quelli di sigarette, qual'è la probabilità che un fiammifero sia stato usato da un fumatore di sigaretta ?

Soluzione:

Se x è il numero di fiammiferi di un fumatore di sigarette:

$10x =$ eventi favorevoli

$10x + 4 \cdot 5x =$ casi totali. Allora $P = 10/30 = 0.33 = 33\%$.

DEFINIZIONI DI PROBABILITÀ

Nelle definizioni e negli esempi precedenti si è usato il concetto di probabilità come acquisito e di uso corrente. In effetti questo è vero in quasi tutti i campi dell'attività umana, ma dare della probabilità una definizione operativa e soddisfacente non è facile. Nel Calcolo delle Probabilità e nel dibattito sui fondamenti di questa disciplina convivono **tre** definizioni di probabilità:

nella **prima**, detta **classica**, la probabilità è il rapporto tra il numero dei casi favorevoli e il numero totale dei risultati possibili, ammesso che questi siano ugualmente possibili. La probabilità classica deriva dai giochi d'azzardo e si usa in presenza di uno schema di casi.

Questa definizione è criticabile perchè usa il termine "*ugualmente possibile*" che è sinonimo di "*ugualmente probabile*"; viene usato, in modo "circolare", il concetto di probabilità per definire la probabilità stessa. Ancora, l'"*ugualmente possibile*" esclude una vastissima classe di fenomeni in cui gli eventi non hanno la stessa probabilità.

La **seconda** definizione si riferisce alla probabilità **frequentista (o frequentizia)**, che trova giustificazione nell'osservazione sperimentale che, ripetendo molte volte un esperimento nelle stesse condizioni, la frequenza di ogni risultato tende a stabilizzarsi attorno ad un valore costante. Questa definizione trova riscontro nei fenomeni fisici, biologici, economico-sociali ed è definita come

$$\lim_{n \rightarrow \infty} \frac{m}{n}$$

con n numero dei risultati possibili ed m numero dei risultati favorevoli.

Le critiche si riferiscono all'impossibilità di trovarsi nelle stesse condizioni e alla difficoltà di misurare un gran numero di casi (es. in un evento mai osservato o difficile, come lo scoppio della III guerra mondiale o l'esplosione di una centrale nucleare).

La **terza** definizione è quella **soggettiva** e si riferisce al grado di fiducia che un individuo assegna al verificarsi di un evento, essendo nota una serie di informazioni. Si basa sul concetto di scommessa, per cui attribuire probabilità $1/2$ vuol dire essere disposto a pagare 1 lira per riceverne 2 se il risultato è favorevole. Questa probabilità è "*non irragionevole*".

Ovviamente il suo difetto è, per definizione, la soggettività e la conseguente incapacità di misurare la probabilità in modo univoco.

- Tutte queste definizioni lasciano perplessi sulla validità di una disciplina scientifica: tuttavia il calcolo delle probabilità ha superato queste incertezze con l'impostazione assiomatica che si basa, in modo totalmente astratto e formale, su un insieme di regole (*assiomi*) cui deve sottostare la valutazione della probabilità e dai quali viene costruita tutta la teoria. I concetti di probabilità e di esperimento vengono assunti come primitivi (e quindi non definiti).

SOMMA E PRODOTTO DI EVENTI

Finora abbiamo trattato i metodi diretti di calcolo delle probabilità, cioè la probabilità di un evento che appartiene ad uno *schema di casi* *

Questi metodi però sono in genere molto laboriosi e in pratica non vengono usati perchè spesso bisogna calcolare le probabilità di eventi la cui riproduzione sperimentale può essere difficile. Ad esempio, si chiedi di trovare la probabilità di abbattere un aereo in un combattimento aereo. È evidente che il calcolo delle probabilità mediante le frequenze è praticamente impossibile, non solo per il costo (anche di vite umane !), ma anche perchè spesso si chiedono i calcoli su un progetto.

Per questo si utilizzano metodi indiretti, che permettono di trovare le probabilità tramite probabilità conosciute di eventi connessi ai primi. Nell'applicazione di questi metodi si fa uso dei **teoremi fondamentali** della teoria della probabilità. Questi sono il teorema della **somma delle probabilità** e il teorema del **prodotto di probabilità**. A voler essere rigorosi, entrambi i teoremi possono essere dimostrati solo per eventi che appartengono ad uno schema di casi. In tutti gli altri casi devono essere postulati.

- Si chiama SOMMA DI DUE EVENTI A e B l'evento C che consiste nel verificarsi dell'evento A o dell'evento B.
- Si chiama SOMMA DI NUMERO QUALSIASI DI EVENTI l'evento C che consiste nel verificarsi di almeno uno di questi eventi.
- Si chiama PRODOTTO DI DUE EVENTI A e B l'evento C che consiste nel verificarsi simultaneo degli eventi A e B.
- Si chiama PRODOTTO DI NUMERO QUALSIASI DI EVENTI l'evento che consiste nel verificarsi simultaneo di **tutti** questi eventi.

Teorema di addizione delle probabilità

La probabilità della somma di due eventi incompatibili è uguale alla somma delle loro probabilità.

$$P(A + B) = P(A) + P(B) \quad (1)$$

* per prove con risultati simmetrici, i casi formano un sistema esaustivo di risultati equiprobabili e mutualmente incompatibili. Questa situazione si chiama **schema di casi**.

evento N.casi
 A m

N.totale di casi $n = m+k$

B k
 $P(A) = m/n$ $P(B) = k/n$

A e B sono incompatibili ($A+B \implies A$ o B): allora $m + k$ casi sono favorevoli alla loro somma.

$$P(A+B) = (m+k)/n$$

Sostituendo nella (1), si ha:

$$(m+k)/n = m/n + k/n \implies \text{identità}$$

• Se gli eventi sono 3 (A, B, C, con $D = A + B$):

$$P(A+B+C) = P(D+C) = P(D) + P(C) = P(A) + P(B) + P(C).$$

Generalizzando:

$$P\left(\sum_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i)$$

Corollario 1

Se gli eventi formano un *gruppo completo di incompatibili*, la somma delle loro probabilità è uguale all'unità.

$$\sum_{i=1}^n P(A_i) = 1.$$

Essendo un gruppo completo l'evento è certo, per cui

$$P(A_1 + A_2 + \dots + A_n) = 1.$$

Dato che sono incompatibili, applichiamo la somma

$$P(A_1 + A_2 + \dots + A_n) = P(A_1) + P(A_2) + \dots + P(A_n) = \sum_{i=1}^n P(A_i) = 1$$

Corollario 2

La somma delle probabilità degli eventi contrari è uguale all'unità (caso particolare del corollario 1).

$$P(A) + P(\bar{A}) = 1.$$

Importante perchè spesso è più facile calcolare l'evento contrario.

• Si chiamano eventi contrari due eventi incompatibili che formano un gruppo completo.

Es.: A = sporcarsi mangiando un gelato; \bar{A} = non sporcarsi mangiando un gelato.

La somma è valida solo per eventi incompatibili. Se sono compatibili, la probabilità della somma è: $P(A+B) = P(A)+P(B)-P(AB)$.

ESEMPIO:

1) In una lotteria ci sono 1000 biglietti, dei quali

1	vince 500 milioni
10	vincono 100 milioni
50	vincono 20 milioni
100	vincono 5 milioni

ciascuno. Tutti gli altri non vincono niente. Trovare la probabilità che il proprietario di un solo biglietto vinca non meno di 20 milioni.

Soluzione:

Gli eventi sono

A vincere non meno di 20 milioni

A1 vincere 20 milioni $P(A1) = 50/1000 = 0.05$

A2 vincere 100 milioni $P(A2) = 10/1000 = 0.01$

A3 vincere 500 milioni $P(A3) = 1/1000 = 0.001$

Sarà $A = A1 + A2 + A3$ e $P(A) = P(A1) + P(A2) + P(A3) = 0.05+0.01+0.001 = 0.061 = 6.1 \%$.

Teorema di moltiplicazione delle probabilità

Prima di enunciare questo teorema, parliamo del concetto di *eventi indipendenti* e *dipendenti*.

• Si dice che l'evento A è dipendente dall'evento B se la probabilità dell'evento A dipende dal fatto che l'evento B si sia verificato o meno.

ESEMPLI:

1) Lancio di due monete. Gli eventi A=testa sulla prima moneta e B= testa sulla seconda sono *indipendenti* perchè la probabilità di A non dipende dall'essersi verificato o meno di B.

2) Urna con 2 palline bianche e 1 nera. Estrazione di una pallina a testa da parte di due persone. Gli eventi A=1.a persona estrae pallina bianca; B=2.a persona estrae pallina bianca, sono *dipendenti* perchè la probabilità di A è 2/3 in assenza di informazioni su B e diventa 1/2 se è noto che B si è verificato.

La probabilità dell'evento A, calcolata a condizione che l'evento B si sia verificato, si chiama **probabilità condizionata** di A e si indica $P(A|B)$.

Dall'esempio 2), $P(A)=2/3$; $P(A|B)=1/2$

L'indipendenza può essere espressa da $P(A|B) = P(A)$, e la dipendenza da $P(A|B) \neq P(A)$.

Teorema di moltiplicazione

La probabilità del prodotto di due eventi è uguale al prodotto della probabilità di uno degli eventi *per* la probabilità condizionata dell'altro, calcolata a condizione che il primo abbia avuto luogo.

$$P(AB) = P(A)P(B | A) \quad (2)$$

Dimostrazione:

Supponiamo che i casi possibili siano \underline{n} , che \underline{m} casi siano favorevoli all'evento A e \underline{k} casi siano favorevoli a B. A e B **non** sono supposti incompatibili e quindi ci saranno casi favorevoli ad entrambi gli eventi (\underline{l} sia il numero di questi casi).

$$\begin{array}{ccc} m \sim A & k \sim B & \\ 000000000000000000000000 & & \\ & l \sim AB & \\ & n & \end{array}$$

Allora sarà: $P(AB) = l/n$ $P(A) = m/n$

Calcoliamo la probabilità condizionata di B, condizionata dal verificarsi di A [$P(B|A)$]. Se A ha avuto luogo, degli \underline{n} casi possibili, restano solo gli \underline{m} favorevoli ad A, dei quali \underline{l} favorevoli a B. Quindi $P(B|A) = l/m$. Sostituendo $P(A)$, $P(AB)$, $P(B|A)$ nella (2) si ottiene un'identità, e con questo il teorema è dimostrato. La (2) si può anche scrivere $P(AB) = P(B)P(A|B)$, dato che l'ordine con cui avvengono A e B non ha importanza.

Corollario 1

Se A è indipendente da B, anche B è indipendente da A.

Dimostrazione: È data la condizione $P(A)=P(A|B)$ [A indipendente da B]; si chiede di dimostrare che $P(B)=P(B|A)$ [B indipendente da A] con $P(A) \neq 0$. È:

$$P(AB) = P(A)P(B | A)$$

$$P(AB) = P(B)P(A | B)$$

da cui $P(A)P(B|A)=P(B)P(A|B)$ o, sostituendo la condizione di indipendenza di A da B, $P(A)P(B|A)=P(B)P(A)$, c.v.d.

Corollario 2

La probabilità del prodotto di due eventi indipendenti è uguale al prodotto delle probabilità degli eventi.

Questo corollario segue immediatamente dalla definizione di evento indipendente.

$$P(AB) = P(A)P(B | A)$$

ma $P(B|A)=P(B)$ se A e B sono indipendenti. Allora $P(AB)=P(A)P(B)$.

- Il corollario si può generalizzare per un numero qualsiasi di eventi. La probabilità del prodotto di un numero qualsiasi di eventi è uguale al prodotto delle probabilità di questi eventi, a condizione che tutti gli $i - 1$ eventi si siano verificati.

$$P(A_1 A_2 \dots A_n) = P(A_1)P(A_2 | A_1)P(A_3 | A_1 A_2) \dots P(A_n | A_1 A_2 \dots A_{n-1})$$

Se gli eventi sono indipendenti, le probabilità non sono più condizionate e l'espressione precedente diventa

$$P(A_1 A_2 \dots A_n) = P(A_1)P(A_2)P(A_3) \dots P(A_n)$$

La probabilità del prodotto di più eventi indipendenti è uguale al prodotto delle probabilità di questi eventi.

- Se gli eventi sono *incompatibili*, non possono essere indipendenti: infatti, dati due eventi incompatibili A e B, se si verifica B, A non può verificarsi e $P(A | B) = 0$, indipendentemente dal valore prevedibile per $P(A)$. Allora $P(AB) = P(B)P(A | B) = 0$ *sempre* e la condizione di indipendenza $P(AB) = P(A) P(B)$ non si verifica mai.

Probabilità totale

La formula della probabilità totale è un corollario dei due teoremi precedenti.

- Si deve determinare la probabilità di un certo evento A che può verificarsi insieme ad uno degli eventi H_1, H_2, H_n , che formano un *gruppo completo di eventi incompatibili*. Questo gruppo viene detto ipotesi. Si ha:

$$P(A) = \sum_{i=1}^n P(H_i)P(A | H_i)$$

Cioè, la probabilità di A è la somma dei prodotti delle probabilità di ciascuna delle ipotesi *per* la probabilità dell'evento con questa ipotesi. [**non la dimostriamo**].

ESEMPI:

- 1) Siano date 3 urne identiche: la prima contiene 2 palline bianche e 1 nera; la seconda 3 palline bianche e 1 nera; la terza 2 palline bianche e 2 nere. Si sceglie a caso un'urna e da essa si estrae una pallina a caso. Trovare la probabilità che la pallina estratta sia bianca.

Soluzione:

	b	n
1	2	1
2	3	1
3	2	2

le ipotesi sono:

H_1 si sceglie la prima urna

H_2 si sceglie la seconda urna

H_3 si sceglie la terza urna

L'evento A consista nell'estrazione di una pallina bianca. Le ipotesi sono equiprobabili e quindi

$$P(H_1) = P(H_2) = P(H_3) = 1/3$$

Le probabilità condizionate di A, con le ipotesi, sono

$$P(A | H_1) = 2/3, P(A | H_2) = 3/4, P(A | H_3) = 1/2$$

$$\begin{aligned} P(A) &= P(H_1) \cdot P(A | H_1) + P(H_2) \cdot P(A | H_2) + P(H_n) \cdot P(A | H_n) \\ &= \frac{1}{3} \frac{2}{3} + \frac{1}{3} \frac{3}{4} + \frac{1}{3} \frac{1}{2} = \frac{2}{9} + \frac{3}{12} + \frac{1}{6} = \frac{23}{36} \end{aligned}$$

2) Ci siano 12 apparecchi in funzione. Di questi, 3 sono costruiti nello stabilimento (1), 4 nello stabilimento (2) e 5 nello stabilimento (3). Gli apparecchi di (1) superano il collaudo con probabilità 0.9; quelli di (2) con probabilità 0.8 e quelli di (3) con 0.75. Trovare la probabilità che un apparecchio preso a caso superi il collaudo.

Soluzione:

Le ipotesi sono:

H_1 l'apparecchio viene da (1)

H_2 l'apparecchio viene da (2)

H_3 l'apparecchio viene da (3)

Le loro probabilità sono:

$$P(H_1) = 3/12=1/4$$

$$P(H_2) = 4/12=1/3$$

$$P(H_3) = 5/12$$

L'evento A consiste in "l'apparecchio ha superato il collaudo" e:

$$P(A | H_1) = 0.9$$

$$P(A | H_2) = 0.8$$

$$P(A | H_3) = 0.75$$

e quindi:

$$P(A) = 1/4 \cdot 0.9 + 1/3 \cdot 0.8 + 5/12 \cdot 0.75 = 0.804$$

Teorema delle ipotesi (formula di Bayes)

Il teorema delle ipotesi (o formula di Bayes) è un corollario del teorema di moltiplicazione e della formula della probabilità totale.

Consideriamo il problema:

Sia dato un gruppo completo di ipotesi incompatibili H_1, H_2, \dots, H_n . Prima di eseguire la prova, le probabilità di queste ipotesi sono note e valgono $P(H_1), P(H_2), \dots, P(H_n)$.

Si esegue un esperimento e si verifica l'evento A . Ci si chiede in che modo cambiano le probabilità delle ipotesi, in seguito al verificarsi dell'evento A . In realtà si cerca la probabilità condizionata $P(H_i|A)$ per ogni ipotesi.

Dal teorema di moltiplicazione si ha:

$$P(AH_i) = P(A)P(H_i|A) = P(H_i)P(A|H_i) \quad i = 1, 2, \dots, n$$

eliminando il primo membro:

$$P(A)P(H_i|A) = P(H_i)P(A|H_i) \quad i = 1, 2, \dots, n$$

da cui:

$$P(H_i|A) = \frac{P(H_i)P(A|H_i)}{P(A)} = \frac{P(H_i)P(A|H_i)}{\sum_{i=1}^n P(H_i)P(A|H_i)} \quad i = 1, 2, \dots, n$$

che è detta formula di Bayes.

Il teorema di Bayes può essere considerato come la traduzione formale di un processo di acquisizione di conoscenza, nel quale le opinioni iniziali, relative al fatto che sia una certa causa ad agire, si combinano con le informazioni derivate dal verificarsi di determinati effetti e si modificano di conseguenza, in un continuo confronto con l'esperienza.

Come esempio si pensi ad un medico che si trovi di fronte ad un paziente che presenta sintomi che sono sempre del colera, ma che a volte si accompagnano a qualche caso di tifo (per fortuna del paziente!). Anche se le probabilità relative sono tutte per il colera, il medico concentrerà la sua attenzione sul tifo, almeno in condizioni ambientali normali, perchè in Italia il tifo è più frequente del colera. In simboli, se indichiamo con C il colera, con T il tifo e con S i sintomi, si ha: $P(C) \ll P(T)$; $P(S|C) = 1$; $P(S|T)$ piccola ma tale che $P(C) < P(S|T)P(T)$ e quindi $P(C|S) < P(T|S)$. Se però nella zona fosse in atto un'epidemia di colera, il medico cambierebbe la sua analisi iniziale affermando che il paziente è affetto da colera.

Accenno ai teoremi delle prove ripetute

Si ha spesso a che fare con il caso in cui sono ripetute molte volte la stessa prova o prove analoghe; in questi casi non interessa il risultato delle singole prove, ma il *numero di realizzazioni* dell'evento A nel corso di una serie di prove. Ci si chiede la probabilità di un numero qualsiasi di realizzazioni dell'evento.

Se le prove sono indipendenti, la soluzione è relativamente semplice e presenta due casi: i risultati delle prove hanno tutti la stessa probabilità di apparire oppure hanno probabilità diversa [1) lancio di una moneta; 2) colpire il bersaglio spostando ogni volta il cannone].

- Nel primo caso avremo il **teorema particolare delle prove ripetute**
- Nel secondo caso avremo il **teorema generale delle prove ripetute**

ESEMPIO:

Si sparino 3 colpi indipendenti su un bersaglio. La probabilità di colpire, per ognuno di loro, è p .
Trovare la probabilità che 2 dei 3 colpi sparati vadano a segno.

Soluzione:

Chiamiamo B_2 il colpire il bersaglio con 2 colpi. L'evento B_2 si può realizzare nei 3 modi:

- 1) 1.o e 2.o colpo a centro, 3.o a vuoto
- 1) 1.o a centro, 2.o a vuoto, 3.o a centro
- 1) 1.o colpo a vuoto, 2.o e 3.o a centro

cioè, chiamando A i centri e \bar{A} i vuoti: $B_2 = A_1A_2\bar{A}_3 + A_1\bar{A}_2A_3 + \bar{A}_1A_2A_3$.

$$P(B_2) = pp(1-p) + p(1-p)p + (1-p)pp$$

(sarebbe stato $P(B_2) = p_1p_2(1-p_3)$, ma le probabilità non sono distinguibili).

$$P(B_2) = ppq + ppq + ppq = 3p^2q$$

Generalizzando il risultato dell'esempio, possiamo risolvere il seguente problema: *Siano n prove indipendenti, in ciascuna delle quali l'evento A può verificarsi o meno; la probabilità del verificarsi di A sia p e la probabilità del suo non verificarsi sia $q=1-p$. Vogliamo determinare la probabilità $P_{m,n}$ che l'evento A , in queste n prove si presenti esattamente (e non "almeno") m volte.*

Tralasciando la dimostrazione, si ha che

$$P_{m,n} = C_n^m p^m q^{n-m}$$

Questa formula, che è esattamente quella dello sviluppo binomiale $(q+p)^n$ è detta **legge binomiale**.

— o —

Nel caso in cui le probabilità dell'evento A siano diverse, si applica il **Teorema generale delle prove ripetute** che qui non viene descritto.

— o —

• Spesso è utile calcolare la probabilità con cui l'evento A avverrà almeno m volte su n casi ($R_{m,n}$):

$$R_{m,n} = \sum_{i=m}^n P_{i,n} = 1 - \sum_{i=0}^{m-1} P_{i,n}$$

ESEMPIO:

La probabilità di ottenere 12, lanciando 2 dadi è $1/36$. Qual'è la probabilità che si verifichi il 12 almeno una volta, lanciando i dadi 24 volte?

Soluzione:

$$P_{0,24} = C_{24}^0 p^0 q^{24} = \left(\frac{35}{36}\right)^{24} = 0.51.$$

Allora:

$$R_{1,24} = 1 - \sum_{i=0}^0 P_{0,24} = 1 - 0.51 = 0.49$$

La probabilità che 12 esca *esattamente* una volta è:

$$P_{1,24} = C_{24}^1 p^1 q^{23} = \frac{24}{36} \left(\frac{35}{36}\right)^{23} = 0.35$$

3. LEGGI DI DISTRIBUZIONE

Le variabili aleatorie (grandezze che possono assumere, nel corso della prova, valori sconosciuti a priori) possono essere discrete e continue.

Le prime possono assumere solo un insieme numerabile di valori, mentre le altre hanno valori che coprono in modo continuo un intervallo.

Esempi di variabili discrete sono:

- il numero di 'testa' uscito in n lanci di una moneta (0,1,2,...,n)
- la frequenza di apparizione di testa nel caso precedente (0, $\frac{1}{n}$, $\frac{2}{n}$, ..., 1)
- il numero di studenti promossi (0,1,2,...,n)
- il numero di fotoni, misurati su un'immagine stellare, in funzione della distanza dal centro della stella

Esempi di variabili continue sono:

- le coordinate di un corpo in moto, nel caso reale
- il tempo di vita di un satellite

Assumiamo di scrivere le variabili aleatorie con le lettere maiuscole e i loro valori (cioè i valori che esse assumono) con le lettere minuscole corrispondenti. Ad esempio, la variabile aleatoria discreta X potrà assumere i valori x_1, x_2, \dots, x_n : potrà cioè verificarsi uno degli eventi che formano il gruppo completo di eventi incompatibili (x_1, \dots, x_n) . La probabilità totale di questo gruppo è ovviamente l'unità. Sarà quindi

$$\sum_{i=1}^n p_i = 1$$

Questa probabilità totale si **distribuisce** in qualche modo tra tutti i valori.

- Per descrivere una variabile aleatoria, bisogna descrivere il modo di distribuirsi della probabilità, ovvero fornire la probabilità di ognuno degli eventi.

In questo modo si definisce la **legge di distribuzione** della variabile aleatoria. Per farlo si stabilisce, in forma tabulare o grafica, una relazione tra i valori della variabile e la loro probabilità (serie di distribuzione e poligonale di distribuzione).

FUNZIONE DI DISTRIBUZIONE

La legge di distribuzione, per come è stata costruita, si applica solo alle variabili discrete, perchè non è possibile costruire la tabella degli infiniti valori di un variabile continua e la probabilità matematica ($\frac{m}{n}$) di infiniti valori è sempre zero ($n=\infty$).

È tuttavia vero che i *diversi intervalli* in cui si possono raccogliere i valori di una variabile continua *non sono equiprobabili* e che *esiste quindi una distribuzione di probabilità degli intervalli*. Per caratterizzare questa distribuzione è comodo usare la probabilità che l'evento assuma un valore **minore** di un valore assegnato [$P(X < x)$ e **non** $P(X=x)$]. La probabilità dell'evento è una funzione di x ; questa funzione si chiama **Funzione di distribuzione della variabile aleatoria X** e si indica come

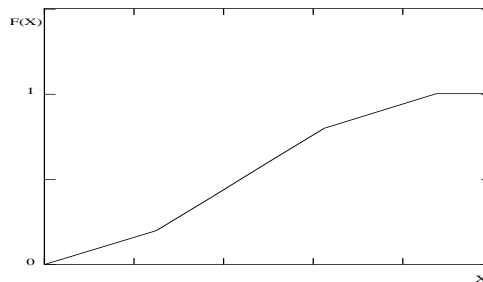
$$F(x) = P(X < x).$$

Si chiama anche *funzione (o legge) integrale di distribuzione*.

La $F(x)$ è la caratteristica più generale di una variabile aleatoria: **esiste** per variabili sia **continue** che **discrete** e caratterizza completamente una variabile aleatoria dal punto di vista probabilistico.

Proprietà generali della $F(x)$

- 1) per $x_2 > x_1$ $F(x_2) \geq F(x_1)$
- 2) per $x \rightarrow -\infty$ $F(x) = 0$ $F(-\infty) = 0$
- 3) per $x \rightarrow +\infty$ $F(x) = 1$ $F(+\infty) = 1$



In generale $F(x)$ è una funzione non decrescente, eventualmente con alcune discontinuità, i cui valori sono compresi tra 0 e 1

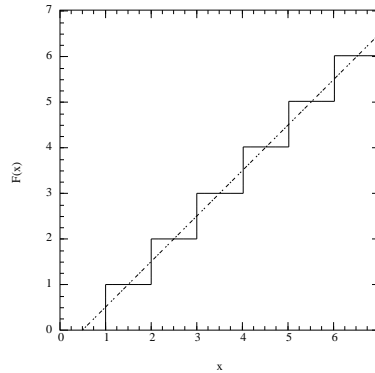
Conoscendo la serie di distribuzione si può facilmente costruire la sua funzione di distribuzione:

$$F(x) = P(X < x) = \sum_{x_i < x} P(X = x_i)$$

ESEMPIO:

- 1) La serie di distribuzione per i possibili valori teorici del lancio di un dado è:

1	1/6	<1	0
2	1/6	<2	1/6
3	1/6	<3	2/6
4	1/6	<4	3/6
5	1/6	<5	4/6
6	1/6	<6	5/6
		<7	6/6



PROBABILITÀ CHE UNA VARIABILE ALEATORIA APPARTENGA AD UN DATO INTERVALLO

Spesso è necessario calcolare la probabilità di avere un valore della variabile aleatoria compreso in un intervallo (α, β) . Questo evento è detto appartenenza della variabile aleatoria X all'intervallo (α, β) .

Si supponga di includere l'estremo sinistro e non l'estremo destro: $\alpha \leq X < \beta$.

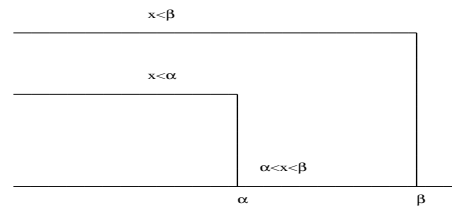
Calcoliamo la probabilità di questo evento (valore di X entro $\alpha \div \beta$), utilizzando la funzione di distribuzione di X. Consideriamo i 3 eventi:

$$A \quad X < \beta$$

$$B \quad X < \alpha$$

$$C \quad \alpha \leq X < \beta$$

con $A = B + C$.



Dal teorema di addizione delle probabilità si ha:

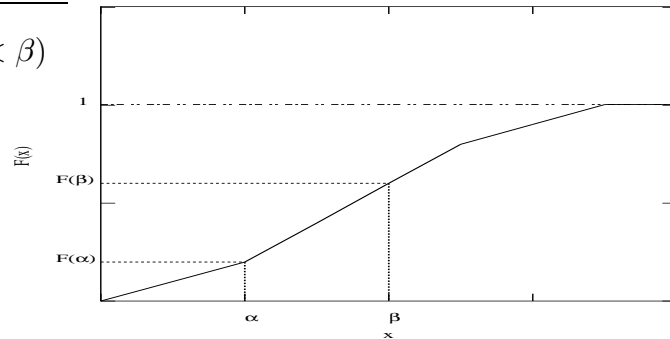
$$P(X < \beta) = P(X < \alpha) + P(\alpha \leq X < \beta)$$

ovvero

$$F(\beta) = F(\alpha) + P(\alpha \leq X < \beta)$$

da cui

$$P(\alpha \leq X < \beta) = F(\beta) - F(\alpha)$$



La probabilità di appartenenza di una variabile aleatoria ad un dato intervallo è uguale all'incremento della funzione di distribuzione nell'intervallo stesso.

Se facciamo decrescere indefinitamente l'intervallo (α, β) , facendo tendere β ad α , otteniamo, come valore limite, la probabilità che la variabile X assuma il valore α .

$$P(X = \alpha) = \lim_{\beta \rightarrow \alpha} P(\alpha \leq X < \beta) = \lim_{\beta \rightarrow \alpha} [F(\beta) - F(\alpha)]$$

Se la $F(x)$, nel punto $X=\alpha$ ha una discontinuità, la $P(X=\alpha)$ è uguale al salto della F . Se $F(x)$ per $x=\alpha$ è continua, il limite sopra è nullo.

D'ora in poi chiameremo variabili aleatorie *continue* quelle variabili la cui funzione di distribuzione F è continua.

Da quanto visto sopra si deduce che:

- La probabilità di un valore qualunque di una variabile aleatoria continua è nulla.

Avevamo visto in precedenza questo fatto, dato che ora i casi possibili sono ∞ .

All'inizio del corso abbiamo incontrato gli eventi impossibili, cioè quelli con probabilità nulla. Ora vediamo che anche gli eventi possibili possono avere probabilità nulla.

Questo concetto è simile a quello di un corpo che ha una certa massa, pur non avendo nessuno dei suoi punti massa finita. Per quanto sia piccolo il volume, un corpo ha massa finita: solo al limite la massa di un punto è nulla. Anche la probabilità, per quanto piccolo possa essere l'intervallo, è diversa da zero. Solo al limite di intervallo nullo la probabilità che un certo valore di X cada in questo intervallo, è zero.

Se la probabilità di un evento è nulla non vuol dire che l'evento non si verificherà mai, cioè *che la sua frequenza è zero*. La frequenza è assimilabile alla probabilità, ma le due grandezze non sono uguali, solo vicine.

$P(X = \alpha) = 0$ significa che in un numero grande a piacere di prove l'evento $x = \alpha$ avverrà raramente quanto si vuole.

Se l'evento A è possibile, ma con $P(A)=0$, l'evento contrario \bar{A} ha probabilità uguale a 1 ($P(\bar{A})=1$), ma non è certo. Con un numero indefinitamente grande di prove l'evento si verificherà **quasi** sempre, ma non sempre.

DENSITÀ DI PROBABILITÀ

Sia X una variabile aleatoria *continua*, associata alla funzione di distribuzione $F(x)$, supposta continua e derivabile. Il calcolo della probabilità per questa variabile di appartenere all'intervallo $(x \div x + \Delta x)$ è

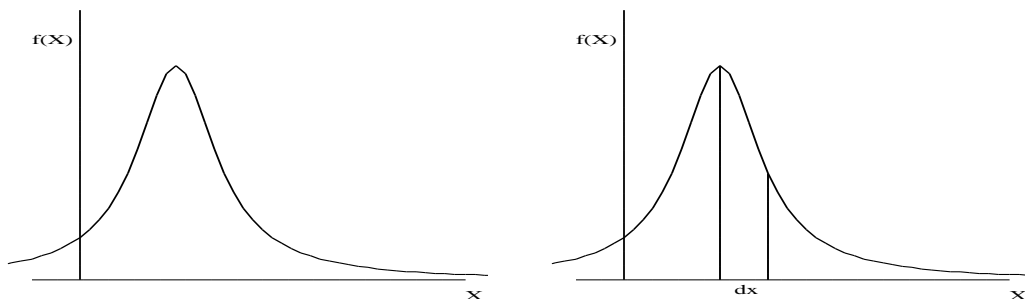
$$P(x < X < x + \Delta x) = F(x + \Delta x) - F(x),$$

cioè, come abbiamo visto, l'incremento della funzione di distribuzione su questo intervallo. Se consideriamo la probabilità media per unità di lunghezza dell'intervallo e facciamo tendere Δx a zero, otteniamo

$$\lim_{\Delta x \rightarrow 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} = F'(x),$$

cioè la derivata prima della funzione di distribuzione. Questa funzione è indicata con il simbolo $f(x)$ e viene detta **densità di probabilità**.

La densità di probabilità, come la funzione di distribuzione, è una delle forme della legge di distribuzione, ma, a differenza di questa, *non è universale* in quanto ha senso solo per variabili aleatorie continue (Fig.1).



Consideriamo (Fig.2) una variabile aleatoria continua X , con densità di probabilità $f(x)$, e l'intervallo elementare dx adiacente al punto x . La probabilità che X appartenga a dx è $f(x)dx$ (geometricamente l'area tratteggiata) è la probabilità elementare. La probabilità di appartenenza di X all'intervallo (α, β) si esprime in funzione della probabilità elementare

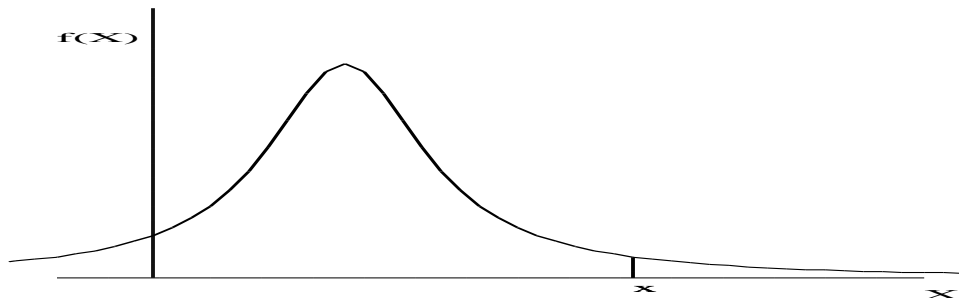
$$P(\alpha < X < \beta) = \int_{\alpha}^{\beta} f(x)dx$$

dove il segno di uguaglianza in α non c'è perchè $P(X=\alpha)=0$.

La $f(x) = F'(x)$ permette di calcolare la densità di probabilità a partire dalla funzione di distribuzione. Il problema inverso, cioè trovare la funzione di distribuzione in funzione della densità è definito da:

$$F(x) = P(X < x) = P(-\infty < X < x) = \int_{-\infty}^x f(x)dx.$$

Geometricamente, $F(x)$ è l'area compresa tra la curva di densità e l'asse delle ascisse, a sinistra del punto x .



Le principali proprietà della densità di probabilità sono:

- 1) È *non negativa* ($f(x) \geq 0$ perchè $F(x)$ è non decrescente)
- 2) $\int_{-\infty}^{+\infty} f(x)dx = 1$ (perchè $F(\infty) = 1$)

ESERCIZIO:

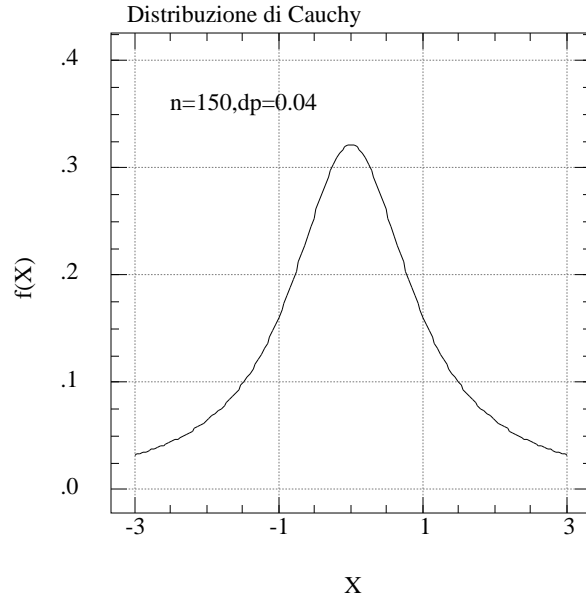
La densità di probabilità della variabile aleatoria X è data da

$$f(x) = \frac{1}{\pi(1+x^2)} \quad (\text{legge di Cauchy})$$

- a) tracciare il grafico della densità
- b) determinare la probabilità che X sia entro (-1,1)

Soluzione:

a)



b)

$$P(-1 < X < 1) = \int_{-1}^1 \frac{dx}{\pi(1+x^2)} = 1/\pi | \arctan x |_{-1}^1 = 1/\pi(\pi/4 + \pi/4) = 1/2$$

CARATTERISTICHE NUMERICHE

Abbiamo visto in precedenza alcune caratteristiche delle variabili aleatorie e cioè le *leggi di distribuzione*. Queste caratteristiche sono:

- per variabili discrete
 - a) funzione di distribuzione
 - b) tabelle di distribuzione (poligono di distribuzione)
- per variabili continue
 - a) funzione di distribuzione
 - b) densità di probabilità (curva di distribuzione)

Ogni legge di distribuzione è sinteticamente rappresentata da una funzione, ed è sufficiente fornire questa funzione per descrivere completamente la variabile aleatoria dal punto di vista probabilistico. In molte applicazioni non è però necessario caratterizzare completamente la variabile aleatoria; è sufficiente fornire alcuni parametri numerici che definiscono in qualche modo alcune delle caratteristiche della variabile. Es.: un valore attorno al quale si distribuiscono i dati; una misura della loro maggiore o minore concentrazione attorno al valore precedente; un numero che definisca il loro grado di simmetria, ecc..

Queste quantità sono dette **caratteristiche numeriche** e sono molto importanti perchè semplificano i calcoli. Esistono molte caratteristiche numeriche, relative ai diversi domini di applicazione; noi vedremo le più comuni.

Valore di aspettazione, Moda, Mediana

Il **valore di aspettazione**, a volte detto **valore medio** o **speranza matematica**, è un numero che rappresenta la variabile aleatoria, almeno in prima approssimazione, e può essere usato al suo posto. Per definizione è:

$$E(x) = \hat{x} = \sum_{i=1}^n x_i P(x_i)$$

per variabili discrete, e

$$E(x) = \hat{x} = \int_A^B x f(x) dx$$

per variabili continue.

• Una qualsiasi funzione di una variabile aleatoria è ancora una variabile aleatoria ed è detta statistica della variabile aleatoria. Ad esempio, se $h(x) = e^{-x}$, diremo che $h(x)$ è una statistica di x .

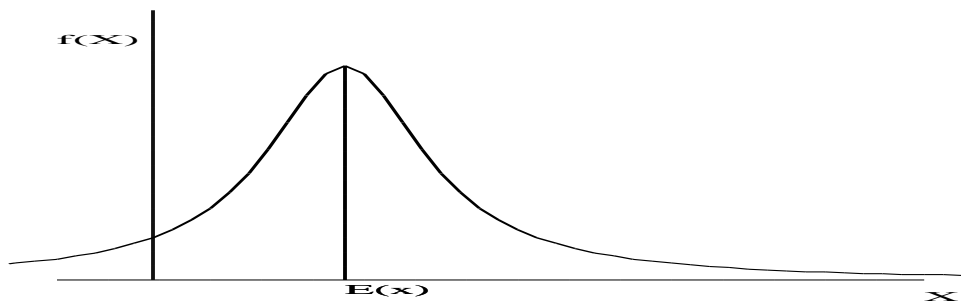
Se $H(x)$ è una statistica di x , il valore di aspettazione di $H(x)$ è:

$$E\{H(x)\} = \sum_{i=1}^n H(x_i) P(x_i)$$

o

$$E\{H(x)\} = \int_A^B H(x) f(x) dx$$

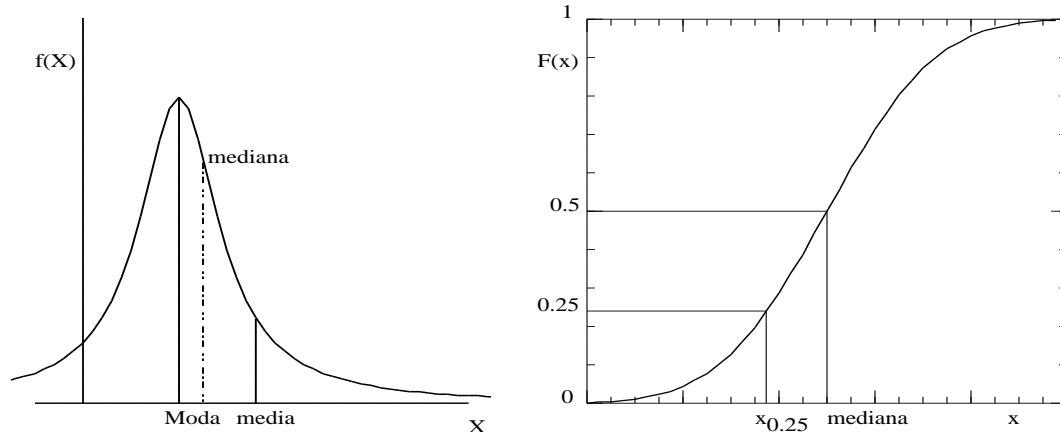
Il valore di aspettazione è legato alla media aritmetica dalla stessa relazione che esiste tra la frequenza e la probabilità: per un gran numero di prove *la media aritmetica tende (converge in probabilità) al valore di aspettazione*. Questa affermazione fa parte della **legge dei grandi numeri** e può esser dimostrata rigorosamente. Il valore di aspettazione così definito corrisponde al centro di gravità di punti materiali di massa diversa.



• Si chiama **moda** di una variabile aleatoria il suo valore più probabile. Nel caso di una variabile continua è il valore per cui la densità di probabilità ammette un massimo. Se la curva (poligono) di distribuzione ha più di un massimo, la distribuzione è detta *multimodale* (se ne ha uno, è detta modale)

• Si chiama **mediana** il valore \underline{m} della variabile X tale che $P(X < m) = P(X > m)$. Geometricamente è l'ascissa del punto che divide in due aree uguali la $f(x)$. Questo si esprime anche con la notazione $F(x_{0.5}) = 0.5$

Se la distribuzione è simmetrica e modale, il valore di aspettazione, la moda e la mediana coincidono.



Si può estendere il concetto di mediana, cioè $F(x_{0.5}) = 0.5$, ad altri valori. Questi prendono il nome di **quantili** (una volta si chiamavano *frattili*) e si definiscono come

$$F(x_q) = \int_{-\infty}^{x_q} f(x) dx = q \quad 0 < q < 1$$

Due valori particolari, $F(x_{0.25}) = 0.25$ e $F(x_{0.75}) = 0.75$ sono detti quantili.

I quantili si possono calcolare dalle tavole della densità di probabilità, utilizzando la definizione data sopra. A titolo di confronto, in fondo al testo (pag.93) viene riportata la tavola dei quantili relativi alla distribuzione normale.

Momenti, Varianza

Si chiama **momento iniziale di ordine l** la quantità

$$m_l = \sum_{i=1}^n x_i^l p(x_i),$$

che corrisponde alla definizione di momento in Meccanica, se le masse p_1, p_2, \dots, p_n sono concentrate lungo l'ascissa, nei punti x_1, x_2, \dots, x_n .

Se la variabile aleatoria è continua, la definizione di momento iniziale di ordine l è:

$$m_l = \int_{-\infty}^{\infty} x^l f(x) dx.$$

Il valore di aspettazione è quindi il momento iniziale di ordine 1.

- Se consideriamo la statistica di X $H(X) = (x - c)^l$, con c numero reale, la quantità

$$\alpha_l = E\{(x - c)^l\}$$

é detta momento di ordine l attorno a c . Se per c viene scelto il valore $E(x) = \hat{x}$, i momenti diventano

$$\mu_l = E\{(x - \hat{x})^l\}$$

e sono detti **momenti attorno al valor medio** (momenti centrali). É $\mu_0 = 1$ e $\mu_1 = 0$. Infatti:

$$\mu_1 = E\{(x - \hat{x})\} = \int_{-\infty}^{\infty} (x - \hat{x}) f(x) dx = \int_{-\infty}^{\infty} x f(x) dx - \int_{-\infty}^{\infty} \hat{x} f(x) dx = \hat{x} - \hat{x} = 0$$

$$\begin{aligned} \mu_2 = E\{(x - \hat{x})^2\} &= E\{x^2 - 2x\hat{x} + \hat{x}^2\} = E(x^2) - 2\hat{x}E(x) + \hat{x}^2 = E(x^2) - 2\hat{x}^2 + \hat{x}^2 \\ &= E(x^2) - \hat{x}^2 \end{aligned}$$

Il momento centrale di ordine 2 viene detto **varianza** ed indicato con $var(x)$ o $\sigma^2(x)$. La sua radice quadrata $\sigma(x)$ viene chiamata **dispersione**, **deviazione standard** o **scarto quadratico medio**.

Proprietá del valore di aspettazione

1) Se \underline{a} e \underline{b} sono due *costanti* e \underline{X} é una variabile aleatoria:

$$E\{aX + b\} = aE\{X\} + b.$$

dimostrazione: per variabili *discrete*

$$E\{aX + b\} = \sum_i (ax_i + b)p_i = a \sum_i x_i p_i + b \sum_i p_i = aE(X) + b$$

per variabili *continue*

$$E\{aX + b\} = \int_{-\infty}^{\infty} (ax + b)f(x)dx = a \int_{-\infty}^{\infty} xf(x)dx + b \int_{-\infty}^{\infty} f(x)dx = aE(X) + b.$$

2) Se X e Y sono due variabili aleatorie

$$E\{X + Y\} = E\{X\} + E\{Y\}$$

Il valore di aspettazione della somma é uguale alla somma dei valori di aspettazione. (*omettiamo la dimostrazione*).

Proprietá della varianza

1) Se \underline{a} e \underline{b} sono due *costanti* e \underline{X} é una variabile aleatoria:

$$var\{aX + b\} = a^2 var\{X\} = a^2 \sigma^2(X).$$

dimostrazione:

$$\begin{aligned} \text{var}\{aX + b\} &= E\{(aX + b - E\{aX + b\})^2\} \\ E\{aX + b\} &= aE(X) + b \end{aligned}$$

allora

$$\begin{aligned} \text{var}\{aX + b\} &= E\{(a^2(X - E\{X\})^2)\} = a^2 E\{(X - E\{X\})^2\} = \\ &= a^2 E\{(X - \hat{X})^2\} = a^2 \text{var}(X) = a^2 \sigma^2(X) \end{aligned}$$

2) Se X e Y sono due variabili aleatorie indipendenti

$$\sigma^2(X + Y) = \sigma^2(X) + \sigma^2(Y)$$

la varianza della somma é uguale alla somma delle varianze

dimostrazione:

$$\begin{aligned} \sigma^2(X + Y) &= E\{[X + Y - E\{X + Y\}]^2\} = E\{[(X - E\{X\}) + (Y - E\{Y\})]^2\} = \\ &= E\{(X - E\{X\})^2 + (Y - E\{Y\})^2 + 2(X - E\{X\})(Y - E\{Y\})\} \end{aligned}$$

Il doppio prodotto é nullo per via dell'indipendenza (covarianza) e quindi:

$$\sigma^2(X + Y) = E\{(X - \hat{X})^2\} + E\{(Y - \hat{Y})^2\} = \sigma^2(X) + \sigma^2(Y)$$

3)

$$\sigma^2(X - Y) = \sigma^2(X) + \sigma^2(Y).$$

Infatti

$$\sigma^2[X + (-Y)] = \sigma^2(X) + \sigma^2(-Y) = \sigma^2(X) + \sigma^2(Y)$$

perché $\sigma^2(aX) = a^2 \sigma^2(X)$.

Variabile standardizzata

Definiamo la statistica di x

$$u = \frac{x - \hat{x}}{\sigma(x)}.$$

I momenti iniziale di primo ordine e centrale di secondo ordine di u sono

$$E(u) = \hat{u} = E\left\{\frac{x - \hat{x}}{\sigma(x)}\right\} = \frac{1}{\sigma(x)} E(x - \hat{x}) = \frac{1}{\sigma(x)} (\hat{x} - \hat{x}) = 0$$

$$\sigma^2(u) = \frac{1}{\sigma^2(x)} E\{(x - \hat{x})^2\} = \frac{\sigma^2(x)}{\sigma^2(x)} = 1$$

da cui si vede che u é una variabile aleatoria che ha media nulla e varianza unitaria. Questa variabile risulta essere molto comoda nei calcoli statistici (viene espressa in unità di σ) e si chiama variabile *ridotta*, *standardizzata*, *normalizzata*, *adimensionale*.

Funzione generatrice dei momenti

Come già visto in precedenza, i momenti sono quantità in grado di individuare completamente una distribuzione. Per questo motivo è importante poter disporre di una **funzione generatrice dei momenti**, che sia in grado di "costruire" i momenti di ogni ordine.

Definizione – data una variabile casuale X e una variabile reale t , si chiama funzione generatrice dei momenti di X , $M(t)$, il valore di aspettazione (*se esiste*) di e^{tX} ; in simboli:

$$M(t) = E\{e^{tX}\} = \sum_x e^{tx_i} p_i$$

$$M(t) = E\{e^{tX}\} = \int_{-\infty}^{\infty} e^{tx} f(x) dx$$

per variabili discrete e continue, rispettivamente.

Quando esiste, $M(t)$ è una funzione di t . Per $t = 0$ $M(t)$ esiste sempre ed è uguale a 1; per gli altri valori di t può non esistere, dato che dipende dalla distribuzione di X .

• La funzione generatrice "genera" tutti i momenti non centrati. Infatti, sviluppando in serie e^{tx} si ha:

$$e^{tX} = 1 + tX + (t^2/2!)X^2 + (t^3/3!)X^3 + \dots + (t^n/n!)X^n \quad ,$$

da cui, prendendo il valore di aspettazione di entrambi i membri, si ha:

$$M(t) = 1 + tm_1 + (t^2/2!)m_2 + (t^3/3!)m_3 + \dots + (t^n/n!)m_n \quad .$$

Da questa, derivando r volte $M(t)$ in $t = 0$, si ottengono i momenti non centrati di ordine r :

$$\frac{d^r}{dt^r} M(t)|_{t=0} = m_r$$

I **momenti centrati** si ricavano sviluppando $E\{(x - \hat{x})^n\}$. Es.: $\mu_2 = E\{x^2\} - \hat{x}^2 = m_2 - m_1^2$; $\mu_3 = m_3 - 3m_1m_2 + 2m_1^3$.

Esempio: Viene mostrato come esempio il calcolo della funzione generatrice di una v.a. continua con un metodo meno veloce dello sviluppo in serie visto in precedenza.

Sia X una variabile casuale con densità di probabilità $f(x) = \lambda e^{-\lambda x}$ per $x \geq 0$. Allora

$$M(t) = E\{e^{tX}\} = \int_{-\infty}^{\infty} e^{tx} \lambda e^{-\lambda x} dx = \lambda \int_{-\infty}^{\infty} e^{-(\lambda-t)x} dx = \frac{\lambda}{\lambda-t} \quad t < \lambda$$

Dato che

$$\frac{d}{dt}M(t) = \frac{\lambda}{(\lambda - t)^2}$$

allora $m_1 = E\{X\} = \frac{d}{dt}M(t)|_{t=0} = \frac{1}{\lambda}$

e che

$$\frac{d^2}{dt^2}M(t) = \frac{2\lambda}{(\lambda - t)^3} ,$$

si ha $m_2 = E\{X^2\} = \frac{d^2}{dt^2}M(t)|_{t=0} = \frac{2}{\lambda^2}$

Il momento centrale del secondo ordine vale:

$$\mu_2 = E\{X^2\} - \hat{X}^2 = 2/\lambda^2 - 1/\lambda^2 = 1/\lambda^2$$

Teorema dell'asse parallelo. Fra tutti i momenti possibili, si preferiscono quelli calcolati rispetto al valore di aspettazione (centrali) perché assumono un minimo. Infatti, riscrivendo e sviluppando per $l = 2$ il momento attorno a c :

$$\begin{aligned} \alpha_2 &= E\{(x - c)^2\} = \sum_{i=1}^n (x_i - c)^2 p_i = \sum_{i=1}^n (x_i - \hat{x} + \hat{x} - c)^2 p_i \\ &= \sum_{i=1}^n (x_i - \hat{x})^2 p_i + 2(\hat{x} - c) \sum_{i=1}^n (x_i - \hat{x}) p_i + (\hat{x} - c)^2 = \mu_2 + (\hat{x} - c)^2 \end{aligned}$$

che ha un minimo per $c = \hat{x}$, cioè quando il momento è calcolato rispetto al valore di aspettazione.

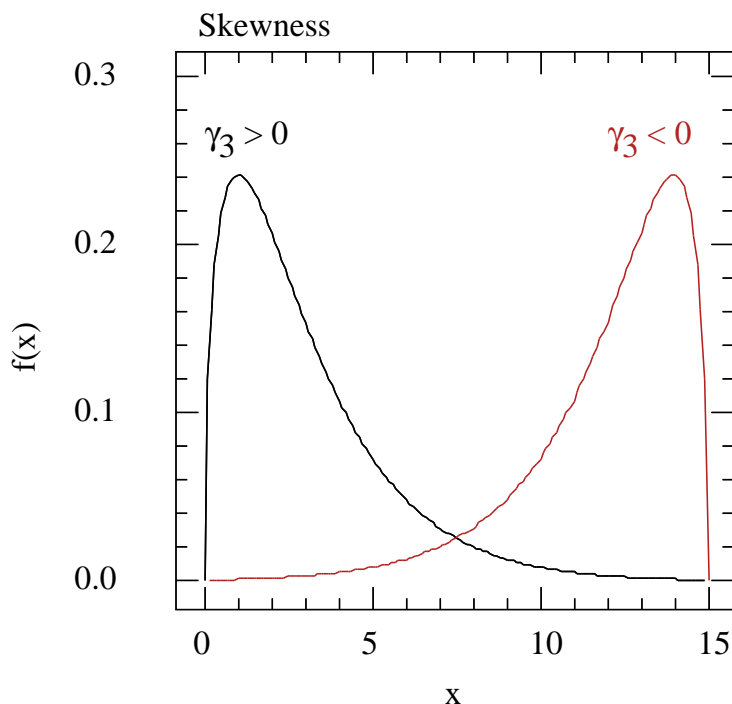
(• da notare che per $c=0$ si trova $\alpha_2 = m_2 = \mu_2 + \hat{x}^2$.)

La stessa cosa succede per i momenti di ordine più elevato. Gli altri momenti che usiamo sono:

$$\begin{aligned} \mu_3 &= E\{(x - \hat{x})^3\} = \sum_{i=1}^n (x_i - \hat{x})^3 p_i = E\{x^3 - 3x^2\hat{x} + 3x\hat{x}^2 - \hat{x}^3\} \\ &= m_3 - 3\hat{x}E(x^2) + 3\hat{x}^2E(x) - \hat{x}^3 = m_3 - 3\hat{x}m_2 + 3\hat{x}^3 - \hat{x}^3 = m_3 - 3m_2\hat{x} + 2\hat{x}^3 \end{aligned}$$

$$\mu_4 = E\{(x - \hat{x})^4\} = \sum_{i=1}^n (x_i - \hat{x})^4 p_i = \dots = -3x^4 + 6\hat{x}^2m_2 - 4\hat{x}m_3 + m_4$$

I momenti centrali descrivono in qualche modo la distribuzione dei singoli valori della variabile aleatoria rispetto al valore di aspettazione. In particolare bisogna notare che i momenti dispari hanno valori nulli se la distribuzione è simmetrica (perché le potenze dispari sono positive e negative). È allora naturale attribuire ai **momenti centrali dispari** una misura della **simmetria della distribuzione**. Il più semplice è il momento centrale di ordine 3 che, usato nella forma adimensionale $\gamma_3 = \mu_3/\sigma^3$ con il nome di coefficiente di asimmetria o **SKEWNESS**, ci fornisce una **asimmetria positiva se $\gamma_3 > 0$** (più asimmetrica a destra) o **negativa se $\gamma_3 < 0$** .



Il momento centrale di ordine 4 fornisce una misura dell'appiattimento o meno della distribuzione. La sua forma adimensionale $\gamma_4 = \mu_4/\sigma^4$ assume il valore 3 per una legge normale; è quindi naturale (?) assumere $\gamma_4 = \mu_4/\sigma^4 - 3$. Questo coefficiente di appiattimento, o KURTOSIS, è uguale a zero se la distribuzione è normale; è positivo se la distribuzione è più acuta di quella normale (leptocurtica); è negativo se la distribuzione è più schiacciata di quella normale (platicurtica).

Ricapitolando:

- Valore di aspettazione - momento iniziale di ordine 1
- Varianza - momento centrale di ordine 2
- Skewness - momento centrale di ordine 3 (adimensionale)
- Kurtosis - momento centrale di ordine 4 (adimensionale)

Disuguaglianza di Chebyshev

La disuguaglianza di Chebyshev è un utile strumento teorico che rappresenta una relazione tra la varianza e la nozione intuitiva di dispersione in una distribuzione. In parole povere ci dice che *maggiori sono le deviazioni dalla media, minore è la probabilità che avvengano* (le maggiori deviazioni accadono di rado). Si utilizza per trovare un **limite superiore** della probabilità di ottenere un valore della variabile aleatoria, quando non si conosce la densità di probabilità. Formalmente, la disuguaglianza si esprime nei tre modi equivalenti:

- 1) $P(|x - \hat{x}| \geq k\sigma) \leq 1/k^2$
- 2) per ogni c costante positiva [$k\sigma = c$]: $P(|x - \hat{x}| \geq c) \leq \sigma_x^2/c^2$
- 3) $P(|x - \hat{x}| < c) \geq 1 - \sigma^2/c^2$

Dimostrazione:

Se chiamiamo $t = (x - \hat{x})^2$ e $P = P(|x - \hat{x}| > k\sigma) \equiv P((x - \hat{x})^2 > k^2\sigma^2)$ si ha $P = \int_{k^2\sigma^2}^{\infty} f(t)dt$ e poiché

$$\sigma^2 = E\{(x - \hat{x})^2\} = \int_{-\infty}^{\infty} t f(t) dt = \int_0^{k^2\sigma^2} t f(t) dt + \int_{k^2\sigma^2}^{\infty} t f(t) dt$$

dato che t é definita positiva e $f(t) \leq 1$ (dens. prob.), si ha, assumendo di poter sostituire alla variabile t il valore dell'estremo inferiore di integrazione (cioé, rispettivamente zero e $k^2\sigma^2$) per ottenere un valore dell'integrale sicuramente minore di quello vero:

$$\sigma^2 \geq 0 \int_0^{k^2\sigma^2} f(t) dt + k^2\sigma^2 \int_{k^2\sigma^2}^{\infty} f(t) dt = k^2\sigma^2 \int_{k^2\sigma^2}^{\infty} f(t) dt = k^2\sigma^2 P$$

cioé

$$\sigma^2 \geq k^2\sigma^2 P((x - \hat{x})^2 \geq k^2\sigma^2) \implies P((x - \hat{x})^2 \geq k^2\sigma^2) \leq \sigma^2/k^2\sigma^2$$

$\sigma = 0$ implica che $P(x = \hat{x}) = 1$. Non la dimostriamo

NOTA: La disuguaglianza di Chebyshev è stata pubblicata nel 1867 dallo stesso P.L. Chebyshev ed è quindi nota con il suo nome. In realtà, però, è stata presentata per la prima volta da I.J. Bienaymé nel 1853. Sarebbe quindi più opportuno chiamarla disuguaglianza di Bienaymé - Chebyshev.

ESERCIZIO:

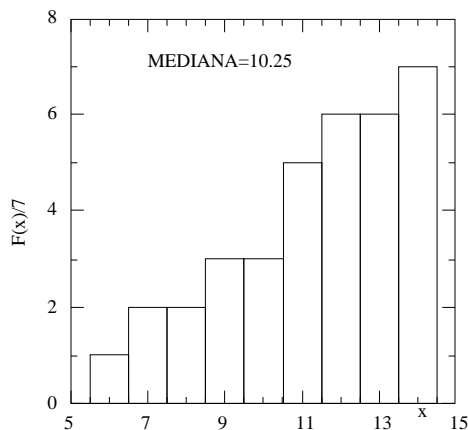
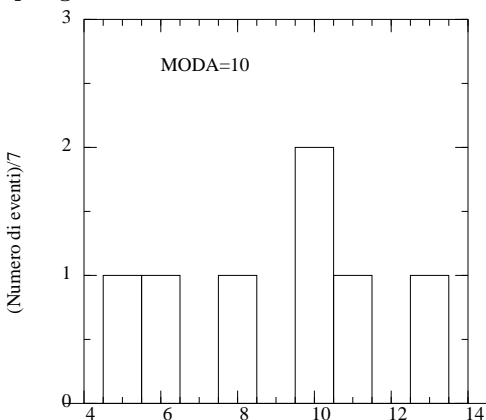
Trovare media, varianza, deviazione standard, skewness, kurtosis, moda e mediana del seguente gruppo di dati:

5,10,11,6,13,10,8

Soluzione:

a)

Il poligono di distribuzione é:



La funzione di distribuzione é: $F(5)=0$; $F(6)=1/7$; $F(7)=2/7$; $F(8)=2/7$; $F(9)=3/7$; $F(10)=3/7$; $F(11)=5/7$; $F(12)=6/7$; $F(13)=6/7$; $F(14)=7/7$. MEDIA - usiamo la definizione di momento iniziale di ordine 1. Chiamiamo \bar{x} la stima di \hat{x} .

$$\bar{x} = Ex = \sum_{i=1}^n x_i p_i = 5/7 + 6/7 + 8/7 + 20/7 + 11/7 + 13/7 = 63/7 = 9 = m_1$$

VARIANZA - usiamo il momento centrale di ordine 2 (i vincoli sono (n-1) perché abbiamo un campione).

$$\mu_2 = E(x - \bar{x})^2 = \sum_{i=1}^n (x_i - \bar{x})^2 p_i = 16/6 + 9/6 + 1/6 + 2/6 + 4/6 + 16/6 = 48/6 = 8 = \sigma^2$$

DEVIATION STANDARD - $\sigma = \sqrt{\sigma^2} = 2.83$

SKEWNESS - Si calcola μ_3 e poi $\gamma_3 = \mu_3/\sigma^3$

$$\mu_3 = E(x - \bar{x})^3 = \sum_{i=1}^n (x_i - \bar{x})^3 p_i = -64/6 - 27/6 - 1/6 + 2/6 + 8/6 + 64/6 = -18/6 = -3$$

$$\gamma_3 = -3/2.83^3 = -3/22.63 = -0.13 \quad (\text{asimmetria a sinistra})$$

KURTOSIS - Prima μ_4 e poi γ_4 .

$$\mu_4 = E(x - \bar{x})^4 = \sum_{i=1}^n (x_i - \bar{x})^4 p_i = 256/6 + 81/6 + 1/6 + 2/6 + 16/6 + 256/6 = 612/6 = 102$$

$$\gamma_4 = \mu_4/\sigma^4 - 3 = 1.59 - 3 = -1.41 \quad (\text{piú piatta di una gaussiana, platycurtica})$$

MODA = 10 - Il valore piú probabile che qui é approssimato dal valore con maggiore frequenza.

MEDIANA - Valore per cui la funzione di distribuzione vale 0.5. La $F(x) = 0.5 = 3.5/7$. Questo valore si ha per $10 < x < 11$. Allora si interpola linearmente

$$x = x_1 + \frac{x_2 - x_1}{y_2 - y_1}(y - y_1), \text{ con } y = 3.5/7$$

$$x = 10 + \frac{11 - 10}{5/7 - 3/7} \left(\frac{3.5}{7} - \frac{3}{7} \right) = 10 + \frac{7}{2} \cdot \frac{0.5}{7} = 10.25$$

ESERCIZIO:

Data la densità di probabilità $f(x) = \frac{e^{-|x|}}{2}$ della variabile aleatoria X, trovare:

- il valore di aspettazione
- la varianza
- la deviazione standard
- lo skewness (asimmetria)
- il kurtosis (appiattimento)

Soluzione:

a)

$$m_1 = \frac{1}{2} \int_{-\infty}^{\infty} x e^{-|x|} dx = 0 \equiv \mu_1$$

La funzione é simmetrica, quindi la media su x é nulla.

b)

$$\sigma^2(x) = \int_{-\infty}^{\infty} x^2 \frac{e^{-|x|}}{2} dx = 2 \int_0^{\infty} x^2 \frac{e^{-x}}{2} dx = \Gamma(3) = 2! = 2$$

$$\int_0^{\infty} x^n e^{-x} dx = \Gamma(n+1); \quad \Gamma(n+1) = n!$$

c) s.d. = $\sqrt{\sigma^2} = \sqrt{2}$

d) La distribuzione é simmetrica, lo skewness é zero

e)

$$\mu_4 = 2 \int_0^{\infty} x^4 \frac{e^{-x}}{2} dx = 24; \quad \Gamma(5) = 4!$$

$$\gamma_4 = \mu_4/\sigma^4 - 3 = 24/4 - 3 = 3 \quad (\text{piu' acuta, leptocurtica})$$

4. DISTRIBUZIONI DI PIÙ VARIABILI ALEATORIE

(Sistemi di variabili aleatorie)

Molto spesso, nella pratica, si ha a che fare con più di una variabile aleatoria contemporaneamente (es. le coordinate x e y dei fotoni provenienti da un'immagine stellare). Parliamo in questo caso di sistemi di variabili aleatorie, per i quali valgono le stesse regole e si possono calcolare le stesse caratteristiche di una sola variabile aleatoria. Evidentemente la complessità della situazione sarà maggiore e bisognerà aggiungere qualche nuova grandezza alla trattazione precedente.

FUNZIONI DI 2 VARIABILI ALEATORIE

Anche per 2 variabili aleatorie X e Y possiamo scrivere la funzione di distribuzione

$$F(x, y) = P(X < x, Y < y),$$

definita come la probabilità che X sia minore di x e contemporaneamente Y sia minore di y.

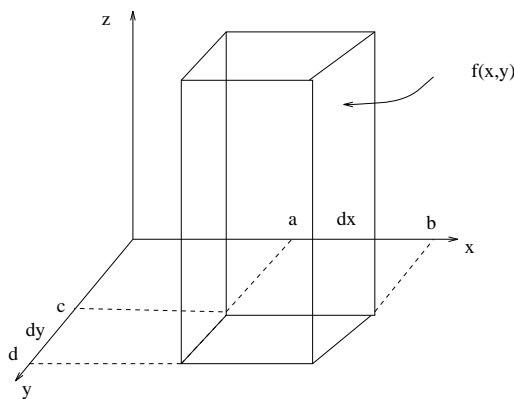
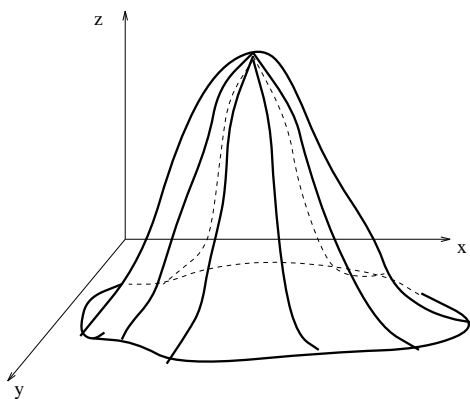
Nel caso monodimensionale era $F(-\infty)=0$ e $F(\infty)=1$: ora è:

$$F(x, -\infty) = F(-\infty, y) = F(-\infty, \infty) = 0$$

$$F(x, \infty) = F_1(x) ; F(\infty, y) = F_2(y) ; F(\infty, \infty) = 1$$

Ripetendo, in due dimensioni, il ragionamento che ci ha portato a definire la densità di probabilità, possiamo scrivere la **densità di probabilità congiunta**:

$$f(x, y) = \frac{\partial}{\partial x} \frac{\partial}{\partial y} F(x, y) = F''(x, y)$$



La probabilità per le variabili x e y di cadere nel rettangolo a, b, c, d è

$$P(a < x < b, c < y < d) = \int_a^b \int_c^d f(x, y) dx dy \quad (1)$$

- La densità di probabilità congiunta è non negativa, e
- il suo integrale :

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1.$$

A volte, della distribuzione bidimensionale interessa solo la variazione in una direzione, indipendentemente dal comportamento nell'altra direzione. In questo caso integriamo la (1) in una dimensione (es. y), su tutti i valori della corrispondente variabile aleatoria:

$$P(a < x < b, -\infty < y < \infty) = \int_a^b \left[\int_{-\infty}^{\infty} f(x, y) dy \right] dx = \int_a^b g(x) dx$$

essendo $g(x) = \int_{-\infty}^{\infty} f(x, y) dy$. La $g(x)$ è chiamata distribuzione marginale di x . La distribuzione marginale di y è $h(y) = \int_{-\infty}^{\infty} f(x, y) dx$.

Le distribuzioni marginali ci permettono di definire anche per le variabili casuali (come già fatto per gli eventi casuali) il concetto di indipendenza. Due variabili aleatorie X e Y sono indipendenti quando

$$f(x, y) = g(x) \cdot h(y),$$

cioè *quando la densità di probabilità congiunta è data dal prodotto delle distribuzioni marginali* delle singole variabili (quando le funzioni sono separabili). Questa condizione è necessaria e sufficiente per l'indipendenza.

Anche in questo caso possiamo parlare di **densità di probabilità condizionata** (in precedenza avevamo introdotto la probabilità condizionata), quando si calcola la

$$P(y \leq Y \leq y + dy | x \leq X \leq x + dx) . \quad (2)$$

Ricordando il teorema di moltiplicazione delle probabilità [$P(AB) = P(A)P(B|A)$], definiamo la densità di probabilità condizionata:

$$f(y|x) = \frac{f(x, y)}{g(x)} \quad \text{oppure} \quad f(y|x) = \frac{g(x) \cdot h(y)}{g(x)} = h(y)$$

(teorema di moltiplicazione delle funzioni di distribuzione).

L'equazione (2) rappresenta la densità di probabilità condizionata elementare $f(y|x)dy$. Un altro modo di enunciare l'indipendenza è: *due variabili aleatorie sono indipendenti quando la conoscenza della densità di probabilità dell'una non modifica la densità di probabilità dell'altra.*

MOMENTI

La definizione di *momento iniziale di ordine l, m* è:

$$m_{l,m} = \sum_{i=1}^n \sum_{j=1}^k x_i^l y_j^m p_{ij} \quad \text{con} \quad p_{ij} = P(X = x_i)P(Y = y_j)$$

per le variabili discrete, e

$$m_{lm} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^l y^m f(x, y) dx dy$$

per quelle continue.

I *momenti attorno ai valori \underline{c} e \underline{d}* sono:

$$\alpha_{lm} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - c)^l (y - d)^m f(x, y) dx dy$$

e quelli *centrali*:

$$\mu_{lm} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \hat{x})^l (y - \hat{y})^m f(x, y) dx dy = E\{(x - \hat{x})^l (y - \hat{y})^m\}.$$

Sapendo che il valore di aspettazione è il momento iniziale di ordine 1, si ha per la funzione di due variabili:

$$E\{xy\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f(x, y) dx dy$$

mentre i valori di aspettazione delle singole variabili sono:

$$m_{10} = E\{x\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy^0 f(x, y) dx dy = \int_{-\infty}^{\infty} xg(x) dx = \hat{x}$$

e

$$m_{01} = E\{y\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^0 y f(x, y) dx dy = \int_{-\infty}^{\infty} yh(y) dy = \hat{y}.$$

Nella pratica si usano soltanto i momenti iniziali di ordine 1 e 2. Per i momenti centrali sono possibili le seguenti combinazioni:

$$\mu_{11} = E\{(x - \hat{x})(y - \hat{y})\} = cov(x, y) \quad 2^{\circ} \text{ ordine misto}$$

$$\mu_{20} = E\{(x - \hat{x})^2\} \quad 2^{\circ} \text{ ordine}$$

$$\mu_{02} = E\{(y - \hat{y})^2\}$$

$$\mu_{00} = m_{00} = 1$$

$$\mu_{10} = \mu_{01} = 0$$

$$m_{10} = \hat{x} = E\{x\}$$

$$m_{01} = \hat{y} = E\{y\}$$

Anche se ora si deve tenere conto sia della \underline{x} che della \underline{y} , abbiamo già visto, nel caso monodimensionale, tutte le grandezze descritte sopra **tranne il momento secondo di ordine misto** μ_{11} .

Questo momento è molto importante ed ha un nome: covarianza di (o tra) x e y . La covarianza rappresenta la dipendenza di una variabile aleatoria dall'altra ed è, in un certo senso, una misura della "covariazione"; indica il grado a cui le variabili sono concordanti o coerenti tra loro.

$$\begin{aligned} cov(x, y) &= E\{(x - \hat{x})(y - \hat{y})\} = E\{xy - x\hat{y} - \hat{x}y + \hat{x}\hat{y}\} = \\ &= E\{xy\} - \hat{y}E\{x\} - \hat{x}E\{y\} + \hat{x}\hat{y} = E\{xy\} - \hat{x}\hat{y} = E\{xy\} - E\{x\}E\{y\} \end{aligned}$$

e, nel caso continuo,

$$\text{cov}(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \hat{x})(y - \hat{y})f(x, y)dx dy.$$

Abbiamo mostrato in precedenza che per *variabili indipendenti* è: $f(x, y) = g(x)h(y)$. Calcoliamo ora la covarianza di due variabili indipendenti:

$$\text{cov}(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \hat{x})(y - \hat{y})f(x, y)dx dy = \int_{-\infty}^{\infty} (x - \hat{x})g(x)dx \int_{-\infty}^{\infty} (y - \hat{y})h(y)dy = 0.$$

essendo **separabile** la densità di probabilità congiunta.

Il primo dei due integrali a destra è il momento centrale di ordine 1 di x (μ_1), che abbiamo visto essere nullo; lo stesso vale per il secondo integrale. Allora, **per due variabili aleatorie indipendenti la covarianza è nulla**. Questo fatto illustra ancora una volta come la covarianza sia una misura della dipendenza esistente tra due variabili aleatorie.

Dalle equazioni di definizione si vede che la covarianza è piccola non solo se la dipendenza è debole, ma anche se una delle distribuzioni (x o y) è poco dispersa (cioè se i valori di x e y si discostano poco dal loro valore di aspettazione). Per eliminare la dipendenza dalla dispersione, alla covarianza si preferisce spesso un **parametro adimensionale** detto **coefficiente di correlazione**, definito come

$$\rho(x, y) = \frac{\text{cov}(x, y)}{\sigma(x)\sigma(y)}$$

Vediamo ora i limiti di variazione di $\rho(x, y)$. Costruiamo le due variabili normali u e v e cerchiamo la varianza della loro somma:

$$\begin{aligned} \sigma^2(u + v) &= E[(u + v) - E(u + v)]^2 = E[(u - \hat{u}) + (v - \hat{v})]^2 = \\ E[(u - \hat{u})^2 + (v - \hat{v})^2 + 2(u - \hat{u})(v - \hat{v})] &= \sigma^2(u) + \sigma^2(v) + 2\text{cov}(u, v) = \\ \sigma^2(u) + \sigma^2(v) + 2\rho(u, v)\sigma(u)\sigma(v) \end{aligned}$$

(le varianze sono additive solo con variabili scorrelate).

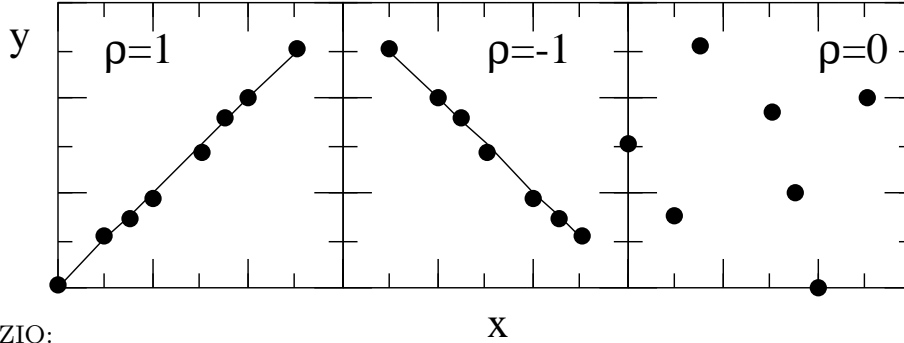
Ma per le variabili standardizzate è $\sigma^2(u) = \sigma^2(v) = 1$ e quindi

$$\begin{aligned} \sigma^2(u + v) &= 2[1 + \rho(u, v)] \\ \sigma^2(u - v) &= 2[1 - \rho(u, v)] \end{aligned}$$

Per avere varianze non negative (sono quadrati) deve essere $-1 \leq \rho(u, v) \leq 1$. Questo è vero anche per $\rho(x, y)$ perché la distribuzione statistica non cambia. Se $\rho = 1$, $\sigma^2(u - v) = 0$ e quindi $u - v = 0$. Allora

$$\begin{aligned} u - v = \frac{x - \hat{x}}{\sigma(x)} - \frac{y - \hat{y}}{\sigma(y)} = 0 \quad (x - \hat{x})\sigma_y - (y - \hat{y})\sigma_x = 0 \\ bx - b\hat{x} - ay + a\hat{y} = 0 \implies y = \frac{a\hat{y} - b\hat{x}}{a} + \frac{bx}{a} = A + Bx \end{aligned}$$

cioè una retta lega x e y , con B positivo (negativo) se ρ è positivo (negativo).



ESERCIZIO:

Calcolare la covarianza della distribuzione discreta di (X,Y) , definita dalle probabilità:

x	
6 8 10	
y 1 :	
2 :	
3 :	

Soluzione

Si usa la formula $cov(x, y) = E(xy) - E(x)E(y)$.

Il valore di aspettazione di X é 8 e quello di Y é 2.

$$E(xy) = 6 \cdot 0.2 + 10 \cdot 0.2 + 16 \cdot 0.2 + 18 \cdot 0.2 + 30 \cdot 0.2 = 16 \implies cov(x, y) = 16 - 16 = 0$$

ESERCIZIO:

Data la densità di probabilità

$$f(x, y) = \begin{cases} 8xy & \text{per } 0 \leq x \leq y \leq 1 \\ 0 & \text{altrimenti} \end{cases}$$

calcolare:

- 1) le distribuzioni marginali $g(x)$ e $h(y)$
- 2) i valori di aspettazione di x e y
- 3) le varianze di x e y
- 4) il coefficiente di correlazione

Soluzione:

1)

$$g(x) = \int_x^1 8xy dy = 4x|y^2|_x^1 = 4x(1 - x^2)$$

$$h(y) = \int_0^y 8xy dx = 4y|x^2|_0^y = 4y^3$$

2)

$$E(x) = \int_0^1 xg(x)dx \quad : \quad E(y) = \int_0^1 yh(y)dy$$

$$E(x) = 4 \int_0^1 x^2(1-x^2)dx = 4 \left[\int_0^1 x^2dx - \int_0^1 x^4dx \right] =$$

$$= 4 [x^3/3|_0^1 - x^5/5|_0^1] = 4/3 - 4/5 = 8/15$$

$$E(y) = 4 \int_0^1 y^4dy = \frac{4}{5}y^5|_0^1 = \frac{4}{5}$$

3)

$$var(x) = \int_0^1 (x - E(x))^2g(x)dx = \int_0^1 [x^2 - 2xE(x) + E^2(x)] 4x(1-x^2)dx =$$

$$= 4 \left[\int_0^1 x^3(1-x^2)dx - 2E(x) \int_0^1 x^2(1-x^2)dx + E^2(x) \int_0^1 x(1-x^2)dx \right] =$$

$$= 4 \left[\frac{1}{4} - \frac{1}{6} - \frac{2}{15}2E(x) + E^2(x)\left(\frac{1}{2} - \frac{1}{4}\right) \right]$$

ma $E(x) = 8/15$. Allora

$$\frac{1}{3} - \frac{16}{15} \frac{8}{15} + \frac{128}{225} - \frac{64}{225} = \frac{1}{3} - \frac{64}{225} = \frac{11}{225}$$

$$var(y) = \int_0^1 (y - E(y))^2h(y)dy = \int_0^1 [y - E(y)]^2 4y^3dy =$$

$$= 4 \left[\int_0^1 (y^5 - 2E(y)y^4 + E^2(y)y^3)dy \right] = \frac{4}{6} - \frac{8}{5}E(y) + E^2(y) = \frac{50 - 48}{75} = \frac{2}{75}$$

4)

Il coefficiente di correlazione è

$$\rho_{x,y} = \frac{E(xy) - E(x)E(y)}{(var x \cdot var y)^{1/2}}$$

$$E(xy) = \int_0^1 \int_0^y xy f(x,y)dxdy = \int_0^1 \left[\int_0^y xy \cdot 8xydx \right] dy = \int_0^1 \left[\int_0^y 8x^2y^2x \right] dy =$$

$$= \int_0^1 8y^2 \frac{x^3}{3} \Big|_0^y dy = \int_0^1 8y^2 \frac{y^3}{3} dy = \frac{8}{3} \int_0^1 y^5 dy = \frac{4}{9}$$

allora:

$$\rho_{x,y} = \frac{\frac{4}{9} - \frac{8}{15} \frac{4}{5}}{\left(\frac{11}{225} \frac{2}{75}\right)^{1/2}} = \frac{12}{75 \cdot 9} \left(\frac{225 \cdot 75}{22}\right)^{1/2} = 0.49$$

FUNZIONI DI N VARIABILI ALEATORIE

Quanto già definito per le funzioni di 1 e 2 variabili può essere esteso al caso generale di n variabili. Quindi la *funzione di distribuzione* sarà:

$$F(x_1, x_2, \dots, x_n) = P(X_1 < x_1, X_2 < x_2, \dots, X_n < x_n).$$

La *densità di probabilità congiunta* (se F è continua e derivabile) è

$$f(x_1, x_2, \dots, x_n) = \frac{\partial^n}{\partial x_1 \partial x_2 \dots \partial x_n} F(x_1, x_2, \dots, x_n).$$

La r-esima distribuzione marginale

$$g_r(x_r) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_{r-1} dx_{r+1} \dots dx_n$$

è la densità di probabilità dell'r-esima variabile aleatoria.

Se definiamo una statistica $H(x_1, x_2, \dots, x_n)$, il *valore di aspettazione* di H è

$$E\{H(x_1, x_2, \dots, x_n)\} =$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} H(x_1, x_2, \dots, x_n) f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n$$

Se $H = x_r$ è

$$E\{x_r\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} x_r f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n$$

cioè

$$E\{x_r\} = \int_{-\infty}^{\infty} x_r g_r(x_r) dx_r$$

Le variabili aleatorie sono *indipendenti* se

$$f(x_1, x_2, \dots, x_n) = g_1(x_1)g_2(x_2) \dots g_n(x_n)$$

Quando le variabili sono due, distribuzione marginale di una è l'integrale dell'altra. Nel caso di $n > 2$ bisogna definire una distribuzione marginale congiunta di k delle n variabili

$$g(x_1, x_2, \dots, x_k) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, x_2, \dots, x_n) dx_{k+1} \dots dx_n$$

Ancora una volta, queste variabili sono indipendenti se la loro densità di probabilità è data dal prodotto delle distribuzioni marginali delle singole variabili.

MOMENTI

I momenti iniziali di ordine $l_1 l_2 \dots l_n$ sono i valori delle statistiche $H = x_1^{l_1} x_2^{l_2} \dots x_n^{l_n}$ e quindi

$$m_{l_1 l_2 \dots l_n} = E\{x_1^{l_1} x_2^{l_2} \dots x_n^{l_n}\}.$$

I valori di aspettazione sono quindi i i momenti iniziali di ordine 1:

$$m_{1000\dots 0} = E\{x_1\} = \hat{x}_1$$

$$m_{0100\dots 0} = E\{x_2\} = \hat{x}_2$$

$$m_{0000\dots 1} = E\{x_n\} = \hat{x}_n$$

I momenti centrali sono

$$\mu_{l_1 l_2 \dots l_n} = E\{(x_1 - \hat{x}_1)^{l_1} (x_2 - \hat{x}_2)^{l_2} \dots (x_n - \hat{x}_n)^{l_n}\}.$$

e quindi le varianze delle singole variabili sono

$$\mu_{2000\dots 0} = E\{(x_1 - \hat{x}_1)^2\} = \sigma^2(x_1)$$

$$\mu_{0200\dots 0} = E\{(x_2 - \hat{x}_2)^2\} = \sigma^2(x_2)$$

⋮

$$\mu_{0000\dots 2} = E\{(x_n - \hat{x}_n)^2\} = \sigma^2(x_n)$$

I momenti centrali misti di ordine 2, cioè

$$c_{ij} = cov(x_i, x_j) = E\{(x_i - \hat{x}_i)(x_j - \hat{x}_j)\}$$

sono le covarianze tra x_i e x_j . È evidente che per $i = j$

$$c_{ii} = cov(x_i, x_i) = \sigma^2(x_i)$$

La definizione delle caratteristiche di una distribuzione di più variabili aleatorie si semplifica molto se adottiamo la notazione vettoriale, cioè se indichiamo le n variabili come le componenti x_1, x_2, \dots, x_n del vettore \mathbf{X} .

La funzione di distribuzione si scrive allora:

$$F = F(\mathbf{X})$$

e la densità di probabilità:

$$f(\mathbf{X}) = \frac{\partial}{\partial \mathbf{X}} F(\mathbf{X}).$$

Il valore di aspettazione della statistica $H(\mathbf{X})$ è:

$$E\{H(\mathbf{X})\} = \int H(\mathbf{X})f(\mathbf{X})d\mathbf{X}.$$

Adesso è facile organizzare le varianze e le covarianze in una matrice, detta **matrice di covarianza** (o varianza-covarianza).

$$C = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & & & \\ c_{m1} & c_{m2} & \cdots & c_{mn} \end{pmatrix}$$

con gli elementi c_{ij} definiti sopra e gli elementi diagonali $c_{ii} = \sigma^2(x_i)$. Come appare dalla definizione di covarianza, $c_{12} = c_{21}$. La matrice di covarianza è simmetrica rispetto alla diagonale principale.

Se interpretiamo anche i valori di aspettazione come un vettore a n -dimensioni, cioè $E\{\mathbf{X}\} = \hat{\mathbf{X}}$, la matrice di covarianza si può scrivere

$$C = E\{(\mathbf{X} - \hat{\mathbf{X}})(\mathbf{X} - \hat{\mathbf{X}})^T\}$$

cioè come il prodotto (diadico) tra il vettore $\mathbf{X} - \hat{\mathbf{X}}$ e il vettore $(\mathbf{X} - \hat{\mathbf{X}})^T$.

ESEMPIO:

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} (b_1 \ b_2 \ \dots \ b_n) = \begin{pmatrix} a_1 b_1 & a_1 b_2 & \cdots & a_1 b_n \\ a_2 b_1 & a_2 b_2 & \cdots & a_2 b_n \\ \vdots & & & \\ a_n b_1 & a_n b_2 & \cdots & a_n b_n \end{pmatrix}$$

5. TRASFORMAZIONI DI VARIABILI

Essendo ormai assodato che una statistica Y della variabile aleatoria \mathbf{X} è anch'essa una variabile aleatoria [$y = y(x)$], ci chiediamo: qual'è la densità di probabilità di Y , $g(y)$, se la densità di X è nota e vale $f(x)$?

ESEMPIO:

La variabile X abbia densità di probabilità:

$$f(x) = 2x \quad 0 < x < 1$$

= 0 *altrove*

Sia A lo spazio $\{x; 0 < x < 1\}$ in cui $f(x) > 0$

Definiamo la variabile aleatoria $Y = 8X^3$ e consideriamo la trasformazione $y = 8x^3$. Questa trasformazione MAPPA lo spazio A nello spazio $B = \{y; 0 < y < 8\}$ ed è inoltre una TRASFORMAZIONE UNO A UNO. Quindi per ogni a e b tali che $0 < a < b < 8$ l'evento $a < Y < b$ avviene quando e solo quando avviene l'evento $\frac{\sqrt[3]{a}}{2} < X < \frac{\sqrt[3]{b}}{2}$. Allora

$$P(a < Y < b) = P\left(\frac{\sqrt[3]{a}}{2} < X < \frac{\sqrt[3]{b}}{2}\right) = \int_{\frac{\sqrt[3]{a}}{2}}^{\frac{\sqrt[3]{b}}{2}} 2x \, dx$$

vogliamo però esprimere l'integrale in termini di y (cerchiamo la densità di probabilità di y). Facciamo la sostituzione: $y = 8x^3 \implies x = \frac{\sqrt[3]{y}}{2}; dx = \frac{dy}{6y^{2/3}}$. I limiti diventano a, b ,

$$P(a < Y < b) = \int_a^b \frac{2\sqrt[3]{y}}{2} \frac{1}{6y^{2/3}} dy = \int_a^b \frac{1}{6y^{1/3}} dy$$

e quindi $g(y) = \frac{1}{6y^{1/3}}$

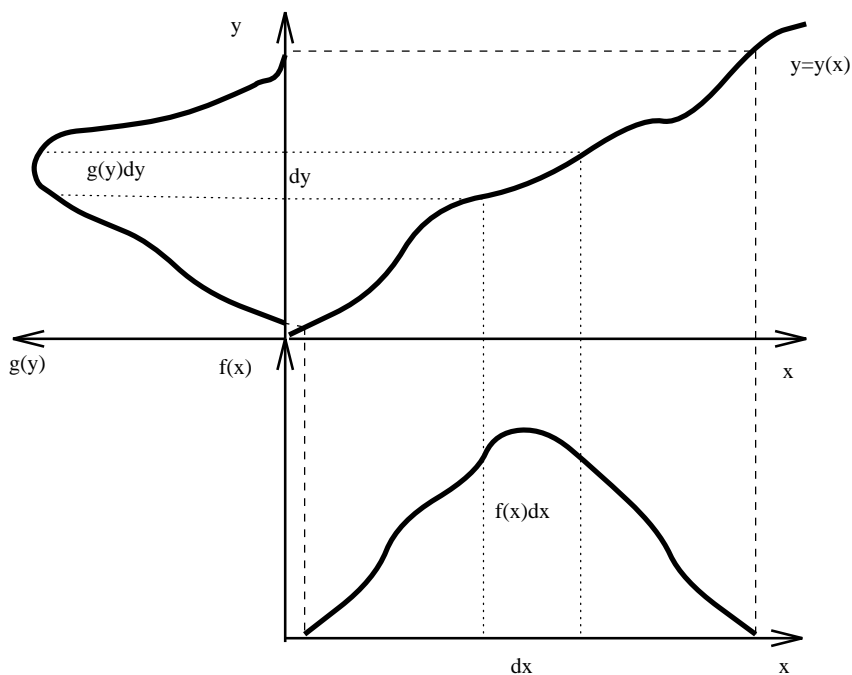
Dalla figura successiva, si vede che, per poter ottenere come $g(y)$ una densità di probabilità (positiva e il cui integrale sia uguale a 1), dovrà essere $f(x)dx = g(y)dy$ e quindi $g(y) = \frac{dx}{dy} f(x)$.

$f(x)$ e $g(y)$ sono non negative, per cui è opportuno usare il valore assoluto di dx e dy . Allora:

$$g(y) = \left| \frac{dx}{dy} \right| f(x)$$

e

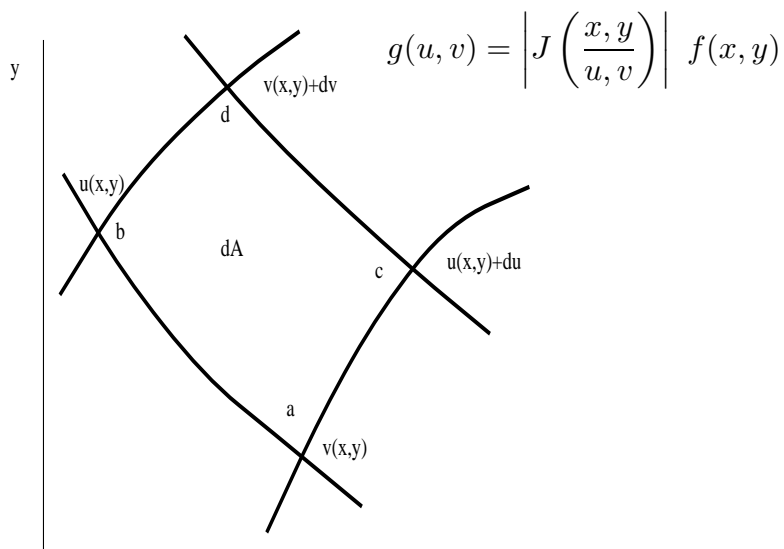
$$\int_{-\infty}^{\infty} g(y) d(y) = \int_{-\infty}^{\infty} f(x) d(x) = 1$$



Questo nel caso di una variabile. Quando abbiamo più variabili, la situazione si complica, ma resta concettualmente la stessa.

Dovremo trovare una funzione J che metta in relazione (connetta) la funzione delle variabili aleatorie e la sua statistica [J corrisponde a dx/dy del caso precedente].

Ci limitiamo al caso di 2 variabili indipendenti \underline{X} e \underline{Y} con le loro statistiche $U = u(x, y)$ e $V = v(x, y)$. Ci proponiamo di trovare una funzione J tale che



Anche in questo caso dovrà essere x

$$\int \int g(u, v) du dv = \int \int f(x, y) dx dy = 1.$$

[La dimostrazione che segue non é da fare]

Nella figura sopra, il quadrilatero dA corrisponde, nel piano (u,v) , all'area infinitesima $dx dy$ nel piano (x,y) . Essendo dA infinitesimo, possiamo calcolare la sua area (in x,y) considerandolo un parallelogramma con vertici a b c d . Le coordinate di a , b , c sono:

$$x_a = x(u, v) \quad y_a = y(u, v)$$

$$x_b = x(u, v + dv) \quad y_b = y(u, v + dv)$$

$$x_c = x(u + du, v) \quad y_c = y(u + du, v)$$

Le coordinate dei vertici b e c possono essere sviluppate in serie di Taylor:*

$$x_b = x(u, v) + \frac{\partial x}{\partial v} dv \quad y_b = y(u, v) + \frac{\partial y}{\partial v} dv$$

$$x_c = x(u, v) + \frac{\partial x}{\partial u} du \quad y_c = y(u, v) + \frac{\partial y}{\partial u} du$$

A meno di un segno, l'area del parallelogramma è data dal determinante:

$$dA = \begin{vmatrix} 1 & x_a & y_a \\ 1 & x_b & y_b \\ 1 & x_c & y_c \end{vmatrix}$$

cioè:

$$dA = \frac{\partial x}{\partial u} du \frac{\partial y}{\partial v} dv - \frac{\partial y}{\partial u} du \frac{\partial x}{\partial v} dv$$

che, scritto come un determinante del secondo ordine, è l'elemento di area dA che corrisponde all'elemento $dx dy$ nel sistema x,y

$$dA = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} \\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} \end{vmatrix} dudv = J \left(\begin{matrix} x, y \\ u, v \end{matrix} \right) dudv$$

Sarà quindi: $g(u, v) = J \left(\begin{matrix} x, y \\ u, v \end{matrix} \right) f(x, y)$.

Il determinante:

$$J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} \\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} \end{vmatrix}$$

viene chiamato **Jacobiano** della trasformazione (con $u = u(x, y)$ $v = v(x, y)$).

Nel caso di n variabili, adottando la notazione vettoriale, si ha

$$\mathbf{X} = (x_1, x_2 \cdots x_n)$$

e la trasformazione

* $f(x) = f(a) + f'(a)(x-a) + \cdots + f''(a) \frac{(x-a)^2}{2!} + \cdots + f^{(n-1)}(a) \frac{(x-a)^{n-1}}{(n-1)!} + R_n$

$$y_1 = y_1(\mathbf{X}) ; y_2 = y_2(\mathbf{X}); \dots ; y_n = y_n(\mathbf{X})$$

Quindi

$$g(\mathbf{Y}) = J \left(\frac{\mathbf{X}}{\mathbf{Y}} \right) f(\mathbf{X})$$

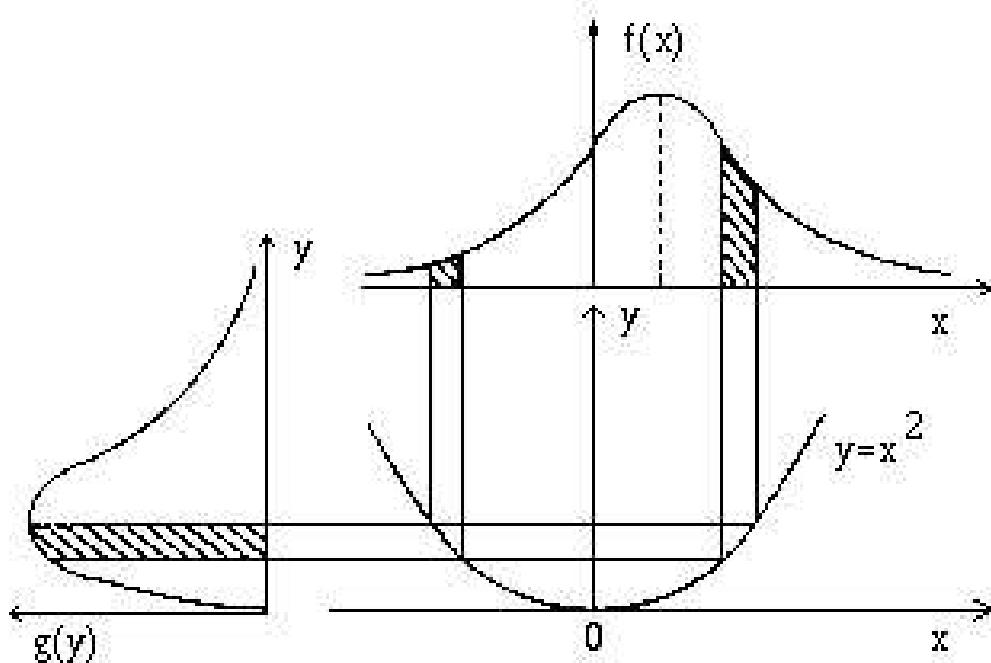
con

$$J \left(\frac{\mathbf{X}}{\mathbf{Y}} \right) = J \left(\frac{x_1, x_2, \dots, x_n}{y_1, y_2, \dots, y_n} \right) = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_2}{\partial y_1} & \dots & \frac{\partial x_n}{\partial y_1} \\ \frac{\partial x_1}{\partial y_2} & \frac{\partial x_2}{\partial y_2} & \dots & \frac{\partial x_n}{\partial y_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_1}{\partial y_n} & \frac{\partial x_2}{\partial y_n} & \dots & \frac{\partial x_n}{\partial y_n} \end{vmatrix}$$

Anche qui, come nel caso monodimensionale, $g(\mathbf{Y})$ esiste se ognuna delle derivate di J è unica.

Esempio di applicazione

Uno dei casi più importanti di applicazione della trasformazione di variabili si ha quando la funzione di trasformazione è del tipo $y = x^2$.



In questo caso si ha:

$$g(y) = \begin{cases} 0 & \text{se } y \leq 0; \\ \frac{f(\sqrt{y}) + f(-\sqrt{y})}{2\sqrt{y}} & \text{se } y > 0 \end{cases}$$

Applichiamo la formula precedente al caso in cui, data una distribuzione normale, si debba trovare la distribuzione della v.a. $y = x^2$. Si ha allora:

$$g(y) = \frac{1}{2\sigma\sqrt{2\pi y}} \left\{ e^{-\frac{(\sqrt{y}-\mu)^2}{2\sigma^2}} + e^{-\frac{(-\sqrt{y}-\mu)^2}{2\sigma^2}} \right\} \quad (5.2)$$

che è un caso particolare della distribuzione χ^2 , descritta più avanti. Se $\mu = 0$ (cioè se la distribuzione normale è centrata) la (5.2) si può scrivere come:

$$g(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{y}{2\sigma^2}} y^{-\frac{1}{2}} \quad (5.3)$$

Con la sostituzione $w = y/\sigma^2 = x^2/\sigma^2$ la (5.3) diventa:

$$h(w) = \frac{1}{\sigma^2\sqrt{2\pi}} e^{-\frac{w}{2}} w^{\frac{1}{2}-1} \quad (5.4)$$

che è una distribuzione χ^2 nel caso in cui ci sia un solo grado di libertà. In particolare la (5.3) è anche la distribuzione del quadrato di una variabile aleatoria normale.

TRASFORMAZIONI LINEARI • PROPAGAZIONE ERRORI

Nella pratica si utilizzano le trasformazioni lineari in quanto sono molto facili da usare. Addirittura trasformazioni di tipo diverso sono spesso approssimate con trasformazioni lineari, usando sviluppi in serie di Taylor.

Le funzioni y_1, y_2, \dots, y_r sono lineari nelle \underline{n} variabili $\mathbf{X} = (x_1, x_2, \dots, x_n)$ quando:

$$y_1 = a_1 + t_{11}x_1 + t_{12}x_2 + \dots + t_{1n}x_n$$

$$y_2 = a_2 + t_{21}x_1 + t_{22}x_2 + \dots + t_{2n}x_n$$

$$y_r = a_r + t_{r1}x_1 + t_{r2}x_2 + \dots + t_{rn}x_n$$

o in forma matriciale : $\mathbf{Y} = T\mathbf{X} + \mathbf{a}$.

Ricordando che è $E\{ax + by\} = aE\{x\} + bE\{y\}$, si ha

$$E\{\mathbf{Y}\} = \hat{\mathbf{Y}} = T\hat{\mathbf{X}} + \mathbf{a}$$

La matrice di covarianza della \mathbf{Y} è

$$\begin{aligned} C_y &= E\{(\mathbf{Y} - \hat{\mathbf{Y}}) (\mathbf{Y} - \hat{\mathbf{Y}})^T\} = \\ &= E\{(T\mathbf{X} + \mathbf{a} - T\hat{\mathbf{X}} - \mathbf{a}) (T\mathbf{X} + \mathbf{a} - T\hat{\mathbf{X}} - \mathbf{a})^T\} \end{aligned}$$

$$E\{T(\mathbf{X} - \hat{\mathbf{X}}) (\mathbf{X} - \hat{\mathbf{X}})^T T^T\} = TE\{(\mathbf{X} - \hat{\mathbf{X}}) (\mathbf{X} - \hat{\mathbf{X}})^T\}T^T$$

e quindi:

$$C_y = TC_x T^T$$

La matrice di covarianza della variabile trasformata dipende dalla matrice di covarianza della variabile originale.

Questo risultato ci porta direttamente alla **legge di propagazione degli errori**.

Si supponga che siano stati misurati i valori di aspettazione \hat{x}_i e che si conoscano gli errori, cioè le deviazioni standard (o le varianze) e le covarianze di \mathbf{X} (e quindi delle x_i). Il nostro scopo è trovare gli errori di una funzione data $y(\mathbf{X})$.

Se gli errori su \mathbf{X} sono piccoli vuol dire che la $f(\mathbf{X})$ è diversa da zero solo in un (piccolo) intorno di \mathbf{X} (pari alla deviazione standard σ). Possiamo quindi sviluppare in serie di Taylor la y in un intorno di $\hat{\mathbf{X}}$:

$$y_i = y_i(\hat{x}) + \left(\frac{\partial y_i}{\partial x_1}\right)_{\mathbf{X}=\hat{\mathbf{X}}}(x_1 - \hat{x}_1) + \dots + \left(\frac{\partial y_i}{\partial x_n}\right)_{\mathbf{X}=\hat{\mathbf{X}}}(x_n - \hat{x}_n) + \dots$$

In notazione matriciale è:

$$\mathbf{Y} = \mathbf{Y}(\hat{\mathbf{X}}) + T(\mathbf{X} - \hat{\mathbf{X}}) + \dots$$

dove

$$T = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \dots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_r}{\partial x_1} & \frac{\partial y_r}{\partial x_2} & \dots & \frac{\partial y_r}{\partial x_n} \end{pmatrix}_{\mathbf{X}=\hat{\mathbf{X}}}$$

Se trascuriamo i termini di ordine superiore nello sviluppo possiamo scrivere la covarianza di y $C_y = TC_x T^T$ come

$$C_y = TE\{(\mathbf{X} - \hat{\mathbf{X}}) (\mathbf{X} - \hat{\mathbf{X}})^T\}T^T$$

Facciamo un esempio 2×2 :

$C_y = TC_x T^T$ diventa:

$$C_y = \begin{vmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} \end{vmatrix} \cdot \begin{vmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{vmatrix} \cdot \begin{vmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} \end{vmatrix} =$$

$$= \begin{vmatrix} \frac{\partial y_1}{\partial x_1} c_{11} + \frac{\partial y_1}{\partial x_2} c_{21} & \frac{\partial y_1}{\partial x_1} c_{12} + \frac{\partial y_1}{\partial x_2} c_{22} \\ \frac{\partial y_2}{\partial x_1} c_{11} + \frac{\partial y_2}{\partial x_2} c_{21} & \frac{\partial y_2}{\partial x_1} c_{12} + \frac{\partial y_2}{\partial x_2} c_{22} \end{vmatrix} \cdot \begin{vmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} \end{vmatrix} =$$

$$= \frac{\partial y_1}{\partial x_1} \left(\frac{\partial y_1}{\partial x_1} c_{11} + \frac{\partial y_1}{\partial x_2} c_{21} \right) + \frac{\partial y_1}{\partial x_2} \left(\frac{\partial y_1}{\partial x_1} c_{12} + \frac{\partial y_1}{\partial x_2} c_{22} \right)$$

Abbiamo calcolato il termine C_{11} della matrice di covarianza di y_1 , cioè la $\sigma^2(y_1)$. Sviluppando l'espressione precedente, si ha:

$$\begin{aligned} C_{11} = \sigma^2(y_1) &= \left(\frac{\partial y_1}{\partial x_1} \right)^2 c_{11} + \frac{\partial y_1}{\partial x_1} \frac{\partial y_1}{\partial x_2} c_{21} + \left(\frac{\partial y_1}{\partial x_2} \right)^2 c_{22} + \frac{\partial y_1}{\partial x_1} \frac{\partial y_1}{\partial x_2} c_{12} = \\ &= \sigma^2(y_1) = \left(\frac{\partial y_1}{\partial x_1} \right)^2 \sigma^2(x_1) + \left(\frac{\partial y_1}{\partial x_2} \right)^2 \sigma^2(x_2) + 2 \frac{\partial y_1}{\partial x_1} \frac{\partial y_1}{\partial x_2} \text{cov}(x_1, x_2) \end{aligned}$$

Se le variabili sono indipendenti ($\text{cov} = 0$), questa espressione è la PROPAGAZIONE DEGLI ERRORI come la conosciamo:

$$\sigma(y_i) = \sqrt{\sum_{j=1}^n \left(\frac{\partial y_i}{\partial x_j} \right)^2 \sigma^2(x_j)}$$

avendo scelto la radice positiva della varianza.

6. ALCUNE FUNZIONI DI DISTRIBUZIONE

Negli esercizi precedenti abbiamo visto alcuni esempi di densità di probabilità. Esistono però alcune distribuzioni che vale la pena di studiare in maggiore dettaglio, per l'importanza (pratica e teorica) che rivestono. E' necessario ricordare che queste funzioni rappresentano una idealizzazione del fenomeno, come la maggior parte dei modelli matematici, che difficilmente può essere messa in pratica; se si preferisce, si può anche dire che è difficile capire se un dato fenomeno segue una legge ideale.

Le densità di probabilità, come abbiamo visto, sono caratterizzate da uno o più parametri, rappresentati dalle caratteristiche numeriche o da una loro funzione.

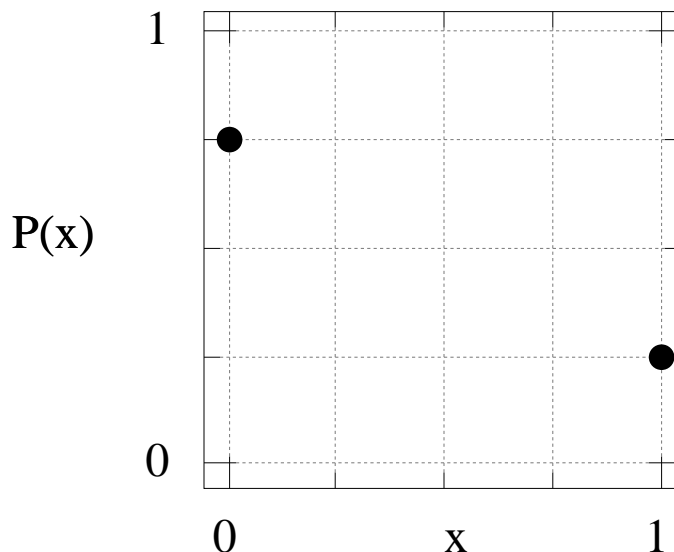
Descriveremo ora alcune densità di probabilità, distinguendo tra quelle relative alle variabili discrete (Bernoulli, Binomiale, Poisson) e quelle relative alle variabili continue (Uniforme, Normale e, successivamente, Student, Fisher, χ^2)

DISTRIBUZIONE DI BERNOULLI

La distribuzione di Bernoulli si applica ogni volta che si considerano due alternative, le cui probabilità restano costanti nel corso di una serie di prove. Queste alternative sono, genericamente, l'evento A e il suo evento contrario \bar{A} , il primo con probabilità $P(A) = p$, il secondo con probabilità $P(\bar{A}) = 1-p = q$. I valori assunti dalla variabile aleatoria di Bernoulli sono, convenzionalmente, 1 per l'evento favorevole e 0 per l'evento contrario ed individuano una dicotomia (si-no; vero-falso; maschio-femmina ...).

Questa distribuzione è caratterizzata dal solo parametro **p**.

La forma della distribuzione di Bernoulli (per $p=0.25$) è:



rappresentata analiticamente da:

$$g(x) = \begin{cases} p^x(1-p)^{1-x} = p^x \cdot q^{1-x} & \text{per } x=0,1 \\ 0 & \text{altrove} \end{cases}$$

con $0 \leq p \leq 1$.

Questa espressione analitica soddisfa le proprietà delle densità di probabilità in quanto **(a)** è sempre non negativa e **(b)** la somma su tutti i valori possibili della v.a. vale 1. Infatti $P(X=0)+P(X=1) = (1-p) + p = 1$.

I momenti

Il valore di aspettazione è

$$E(X) = \sum x_i p_i = 1 \cdot p + 0 \cdot q = p$$

e la varianza

$$\text{var}(X) = \sum (x - \hat{x})^2 p_i = (0 - p)^2 q + (1 - p)^2 p = p^2 q + q^2 p = pq(p + q) = pq.$$

La funzione generatrice dei momenti è:

$$M_X(t) = e^{0 \cdot t} q + e^{t} p = (1 - p) + p \cdot e^t = q + p \cdot e^t$$

DISTRIBUZIONE BINOMIALE

Quando si esaminano n eventi di Bernoulli ci si chiede qual'è la distribuzione di probabilità della variabile aleatoria, $X = \sum_{i=1}^n x_i$ (con x_i variabile di Bernoulli di probabilità p); ci si chiede cioè qual'è la probabilità che l'evento favorevole (di Bernoulli) avvenga m volte in n prove.

È questo il caso cui si è accennato all'inizio del corso, delle *prove ripetute* per le quali la probabilità che la variabile aleatoria X assuma il valore m è

$$P\{X = m\} = W_m^n = \binom{n}{m} p^m q^{n-m}$$

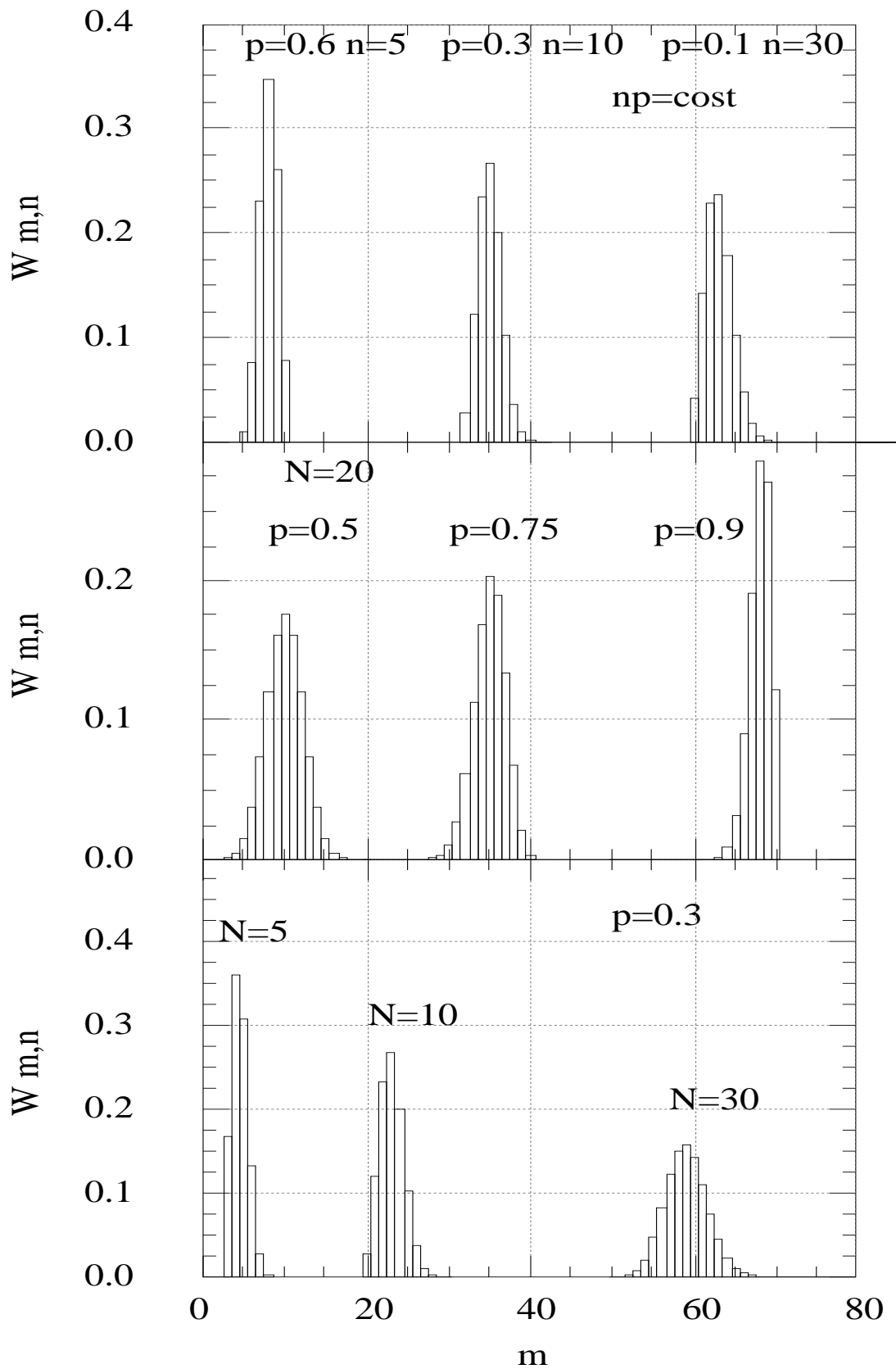
con

$$\binom{n}{m} = \frac{n!}{m!(n-m)!}$$

quindi la forma analitica della binomiale è

$$W_m^n = \frac{n!}{m!(n-m)!} p^m q^{n-m},$$

che rappresenta la probabilità che un evento favorevole avvenga m volte in n tentativi, quando le alternative sono due (maschio-femmina, morto-vivo, accettato-rigettato), se la probabilità che avvenga l'evento è p.



Esempi di distribuzione binomiale al variare dei parametri

CARATTERISTICHE NUMERICHE

$$E(X) = E\left(\sum_{i=1}^n x_i\right) = \sum_{i=1}^n E(x_i) = nE(x) = n(1p + 0q) = np$$

$$\sigma^2(X) = \sigma^2\left(\sum_{i=1}^n x_i\right) = \sum_{i=1}^n \sigma^2(x_i) = n\sigma^2(x_i) = nE\{(x-p)^2\} =$$

$$n[(1-p)^2p + (0-p)^2q] = n[p + p^3 - 2p^2 + p^2 - p^3] =$$

$$= n(p - p^2) = np(1-p) = npq$$

$$\sigma(X) = \sqrt{npq}$$

$$\text{Skewness} = \frac{q-p}{\sqrt{npq}}$$

$$\text{Kurtosis} = \frac{1-6pq}{npq}$$

Funzione generatrice dei momenti: $M_X(t) = (q + pe^t)^n$

I parametri della binomiale sono \underline{n} (numero di dati) e \underline{p} (probabilità del singolo evento).

Studiamo le variazioni di uno o entrambi questi parametri.

- 1) Se $n = \text{cost}$ e p cresce W_m^n diventa più asimmetrica (è simmetrica quando n è grande rispetto a p).
- 2) se $p = \text{cost}$, W_m^n diventa più asimmetrica al diminuire di n (per lo stesso motivo del punto 1).
- 3) Se n cresce e $n \cdot p = \text{cost}$, la distribuzione tende ad un'altra distribuzione che vedremo (Poisson).

ESERCIZIO:

Lanciando un dado 5 volte, si calcoli la probabilità di ottenere il 2 tre volte.

Soluzione:

Tipico caso da binomiale: l'evento favorevole è il 2, quello contrario tutto il resto. Questa è una binomiale con $n = 5$ e $p = \frac{1}{6}$ (nel caso particolare è $m = 3$). Allora la probabilità che $X=3$ è

$$\begin{aligned} P(X = 3) &= \binom{n}{m} p^m q^{n-m} = \frac{n!}{m!(n-m)!} p^m q^{n-m} \\ &= \frac{5!}{3!2!} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^2 = \frac{250}{7776} = 0.032 \end{aligned}$$

Quindi, se facciamo una serie di "misure", ognuna delle quali consistente in 5 lanci, avremo una probabilità del 3.2% di ottenere il 2 tre volte.

ESERCIZIO

La probabilità di laurearsi, per uno studente che si iscrive all' università è 0.3. Trovare la probabilità che di 5 studenti se ne laureino ($n=5$ $p=0.3$):

a) 0, 1, 2, 3, 4, 5

b) almeno 1

a)

$$P(0) = \frac{5!}{0!5!} 0.3^0 \cdot 0.7^5 = 0.168 = 16.8\%$$

$$P(1) = \frac{5!}{1!4!} 0.3^1 \cdot 0.7^4 = 0.36 = 36\%$$

$$P(2) = \frac{5!}{2!3!} 0.3^2 \cdot 0.7^3 = 0.31 = 31\%$$

$$P(3) = \frac{5!}{3!2!} 0.3^3 \cdot 0.7^2 = 0.13 = 13\%$$

$$P(4) = \frac{5!}{4!1!} 0.3^4 \cdot 0.7^1 = 0.028 = 2.8\%$$

$$P(5) = \frac{5!}{5!0!} 0.3^5 \cdot 0.7^0 = 0.0024 = 0.2\%$$

b)

$$P(\text{almeno 1 studente}) = 1 - P(0) = 1 - 0.168 = 0.832 = 83.2\%$$

DISTRIBUZIONE MULTINOMIALE

Se gli eventi A_1, A_2, \dots, A_k avvengono con probabilità p_1, p_2, \dots, p_k , essendo ogni p_j costante e con $\sum_{j=1}^k p_j = 1$, avendo effettuato n esperimenti, la probabilità che A_1, A_2, \dots, A_k avvengano rispettivamente n_1, n_2, \dots, n_k volte è

$$W_{n_1, n_2, \dots, n_k}^n = \frac{n!}{\prod_{j=1}^k n_j!} \prod_{j=1}^k p_j^{n_j} = \frac{n!}{n_1! n_2! \dots n_k!} p_1^{n_1} p_2^{n_2} \dots p_k^{n_k}$$

Questa legge, generalizzazione della binomiale, è detta MULTINOMIALE.

ESERCIZIO:

Lanciando un dado 12 volte, qual'è la probabilità di ottenere 1,2,3,4,5 e 6 esattamente 2 volte?

Soluzione:

$$W_{2,2,2,2,2,2}^{12} = \frac{12!}{2!2!2!2!2!2!} \left(\frac{1}{6}\right)^2 \left(\frac{1}{6}\right)^2 \left(\frac{1}{6}\right)^2 \left(\frac{1}{6}\right)^2 \left(\frac{1}{6}\right)^2 \left(\frac{1}{6}\right)^2 = \frac{1925}{559.872} = 0.00344$$

ESERCIZIO:

Supponiamo che da rilevazioni fatte in anni precedenti, le probabilità che gli studenti che si iscrivono al primo anno di un scuola superiore abbiano ottenuto, all'esame di terza media, i giudizi di "ottimo", "distinto", "buono" e "sufficiente", siano rispettivamente $p_1=0.1$, $p_2=0.25$, $p_3=0.2$ e $p_4=0.45$. Qual'è la probabilità di formare una prima classe di 28 allievi costituita da 4 studenti con "ottimo", 8 con "distinto", 7 con "buono" e 9 con "sufficiente"?

Soluzione:

Si applica la distribuzione multinomiale con parametri $n=28$; p_1, p_2, p_3, p_4 riportati nel testo.

$$W_{4,8,7,9}^{28} = \frac{28!}{4!8!7!9!} (0.1)^4 (0.25)^8 (0.2)^7 (0.45)^9 = 0.25\%$$

Legge dei grandi numeri

Una delle nozioni fondamentali che si usa nella formulazione di un modello di probabilità è quella per cui la probabilità di un evento incorpora il concetto di stabilità della frequenza con cui si verifica l'evento stesso.

Questa proprietà della frequenza, *cioè la possibilità di essere usata al posto della probabilità*, in certe condizioni, prende il nome di **legge dei grandi numeri**.

Nel caso particolare in cui un evento A avviene con probabilità p e non avviene con probabilità $1 - p$, questa legge stabilisce che:

$$\lim_{n \rightarrow \infty} \frac{x_i}{n} = p$$

con x_n numero di eventi favorevoli in n tentativi (la frequenza di A). Quindi si può dire che la frequenza tende alla probabilità quando il numero di prove è alto.

Discutiamo ora una formulazione più rigorosa della legge dei grandi numeri. Questa legge, conseguenza diretta della disuguaglianza di Chebyshev, stabilisce una relazione tra la media aritmetica di un certo campione di una variabile aleatoria ed il valore di aspettazione di questa variabile.

Sia X una variabile casuale, con valore di aspettazione $\hat{x} = E\{X\}$ e varianza $\sigma^2 = var\{X\}$. Si effettuino n determinazioni indipendenti x_1, x_2, \dots, x_n di X ; l'insieme di questi n valori rappresenta n variabili casuali, ciascuna delle quali obbedisce alla distribuzione di X (e quindi possiede \hat{x} e σ^2).

Si consideri la media aritmetica di queste n variabili casuali

$$\bar{X} = \frac{1}{n} \sum x_i.$$

Anche questa media è una variabile casuale, il cui valore di aspettazione è

$$E\{\bar{X}\} = \hat{\bar{x}} = \sum \left(\frac{1}{n} \sum x_i \right) p_i = \frac{1}{n} \sum \hat{x}_i = \frac{1}{n} n \hat{x} = \hat{x}$$

e la cui varianza è:

$$\begin{aligned} var\{\bar{x}\} &= \sigma_{\bar{x}}^2 = \sum_i \left(\frac{1}{n} \sum x_i - \frac{1}{n} \sum \hat{x} \right)^2 p_i = \\ &= \frac{1}{n^2} \sum \left[\sum (x_i - \hat{x})^2 p_i \right] = \frac{1}{n^2} \sum var(x_i) = \frac{1}{n^2} n \sigma^2 = \frac{\sigma^2}{n} \end{aligned}$$

Allora:

- Il valore di aspettazione di \bar{X} non dipende dal numero n delle prove ed è uguale al valore di aspettazione di X .
- La varianza di \bar{X} , invece, decresce al crescere del numero delle prove.

Riprendendo il discorso su probabilità e frequenza, possiamo dire che la frequenza tende alla probabilità con un errore sempre più piccolo al crescere del numero delle prove.

In definitiva, la legge dei grandi numeri può essere enunciata così: *Per un numero sufficientemente grande di prove indipendenti, la media aritmetica dei valori osservati di una variabile aleatoria converge in probabilità al valore di aspettazione di tale variabile.*

- Sia la distribuzione binomiale che la multinomiale richiedono che gli esperimenti siano indipendenti, cioè con rimpiazzamento. Se non c'è rimpiazzamento (ad esempio nei sondaggi) la distribuzione che ne deriva si chiama ipergeometrica (che non facciamo).

DISTRIBUZIONE DI POISSON

Un evento casuale che continua nel tempo (o nello spazio) e che è osservato così come si presenta è chiamato **processo stocastico**, **processo random** o semplicemente **processo**. Un'istantanea, o un'osservazione ad ogni istante di tempo (o in ogni punto dello spazio), è una parte del processo random. La variazione temporale di una variabile numerica del processo costituisce una famiglia di variabili casuali in funzione del tempo. Un processo importante è il cosiddetto **processo di Poisson** che descrive un'ampia varietà di fenomeni, che possiedono certe caratteristiche comuni e nei quali un qualche tipo di avvenimento ha luogo sporadicamente in un dato intervallo di tempo, in modo da poter dire che sia avvenuto casualmente.

Esempi ai quali è possibile applicare il processo di Poisson sono:

- 1) difetti in lungo nastro o filo
- 2) conteggi registrati da un contatore geiger vicino ad una sorgente radioattiva
- 3) difetti in un equipaggiamento o nei componenti di uno strumento
- 4) arrivi di navi in porto, di auto ad un casello autostradale
- 5) rottura dell'auto
- 6) chiamate telefoniche
- 7) la vendita di un apparecchio determinato, in un magazzino
- 8) il numero di stelle per unità di superficie sferica, in cielo

Anche se, come si vede, gli eventi sono riferiti allo spazio, questo processo è convenzionalmente studiato con riferimento al tempo.

Tutti i processi di Poisson soddisfano le seguenti condizioni (postulati di Poisson):

- 1) Gli eventi, definiti come numero di accadimenti in intervalli temporali non sovrapposti, sono indipendenti.
- 2) La forma della distribuzione di probabilità è invariante temporalmente.

3) La probabilità che si verifichi un solo evento in un piccolo intervallo di tempo, è proporzionale alla dimensione dell'intervallo.

4) La probabilità che si verifichi più di un evento in un piccolo intervallo di tempo è trascurabile rispetto alla probabilità che se ne verifichi uno solo.

Gli eventi che seguono questo schema sono descritti da una densità di probabilità che deriva dalla legge binomiale, quando \underline{np} resta *costante* ed \underline{n} aumenta (\underline{p} diminuisce). Se definiamo $\lambda = np$, possiamo scrivere la legge binomiale

$$W_m^n = \frac{n!}{m!(n-m)!} p^m q^{n-m}$$

come

$$\begin{aligned} & \frac{n!}{m!(n-m)!} \left(\frac{\lambda}{n}\right)^m \frac{\left(1 - \frac{\lambda}{n}\right)^n}{\left(1 - \frac{\lambda}{n}\right)^m} = \\ & = \frac{\lambda^m}{m!} \frac{n(n-1)(n-2)\dots(n-m+1)}{n^m} \frac{\left(1 - \frac{\lambda}{n}\right)^n}{\left(1 - \frac{\lambda}{n}\right)^m} = \end{aligned}$$

[per lo sviluppo di $\frac{n!}{(n-m)!}$ vedere Brandt App. C, eq. 1 pag.364]

$$= \frac{\lambda^m}{m!} \left(1 - \frac{\lambda}{n}\right)^n \frac{\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\dots\left(1 - \frac{m-1}{n}\right)}{\left(1 - \frac{\lambda}{n}\right)^m}$$

Nel limite di \underline{n} grande, i termini a destra tendono ad 1, ed essendo

$$\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda} \quad \left[\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n = e \right]$$

si ha:

$$\lim_{n \rightarrow \infty} W_m^n = f(m) = \frac{\lambda^m}{m!} e^{-\lambda}.$$

Questa è la distribuzione di Poisson, definita per valori interi di m . La sua probabilità totale (la probabilità di trovare un qualsiasi valore di m) è 1. Infatti:

$$\sum_{m=0}^{\infty} f(m) = \sum_{m=0}^{\infty} \frac{e^{-\lambda} \lambda^m}{m!} = e^{-\lambda} \left(\frac{\lambda^0}{0!} + \frac{\lambda^1}{1!} + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} \dots \right) = e^{-\lambda} e^{\lambda} = 1$$

essendo il termine tra parentesi lo sviluppo di e^{λ} .

CARATTERISTICHE NUMERICHE

valore di aspettazione

$$E(X) = \sum_{i=0}^{\infty} x_i p(x_i) = \sum_{m=0}^{\infty} m \frac{\lambda^m}{m!} e^{-\lambda} = \sum_{m=1}^{\infty} m \frac{\lambda^m}{m!} e^{-\lambda} =$$

$$\sum_{m=1}^{\infty} \frac{\lambda \lambda^{m-1}}{(m-1)!} e^{-\lambda} = \lambda \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} e^{-\lambda} = \lambda$$

$E(x) = \lambda$: il valore di aspettazione è uguale al parametro.

varianza

$$\begin{aligned} \sigma^2(X) &= E(X^2) - \{E(X)\}^2 \\ E(x^2) &= \sum_{m=1}^{\infty} m^2 \frac{\lambda^m}{m!} e^{-\lambda} = \lambda \sum_{m=1}^{\infty} m \frac{\lambda^{m-1}}{(m-1)!} e^{-\lambda} = \\ &= \lambda \sum_{j=0}^{\infty} (j+1) \frac{\lambda^j}{j!} e^{-\lambda} = \lambda \left(\sum_{j=0}^{\infty} j \frac{\lambda^j}{j!} e^{-\lambda} + 1 \right) \end{aligned}$$

Il termine

$$\sum_{j=0}^{\infty} j \frac{\lambda^j}{j!} e^{-\lambda} = \lambda$$

(v. calcolo del valore di aspettazione). Allora $E(x^2) = \lambda(\lambda + 1)$ e:

$$\sigma^2(X) = E(X^2) - \{E(X)\}^2 = \lambda(\lambda + 1) - \lambda^2 = \lambda$$

$$\sigma^2(x) = \lambda$$

$$\sigma(x) = \sqrt{\lambda}$$

Tutti i momenti valgono λ

Da notare che

$$\text{skewness} = \frac{\mu_3}{\sigma^3} = 1/\sqrt{\lambda} \text{ e}$$

$$\text{kurtosis} = \frac{\mu_4}{\sigma^4} = 1/\lambda \text{ non sono i momenti, ma funzioni dei momenti.}$$

Quando osserviamo un fenomeno in cui la media e la varianza sono uguali, siamo sicuri che gli eventi sono indipendenti.

Funzione generatrice dei momenti

La funzione generatrice dei momenti per la distribuzione di Poisson è data da:

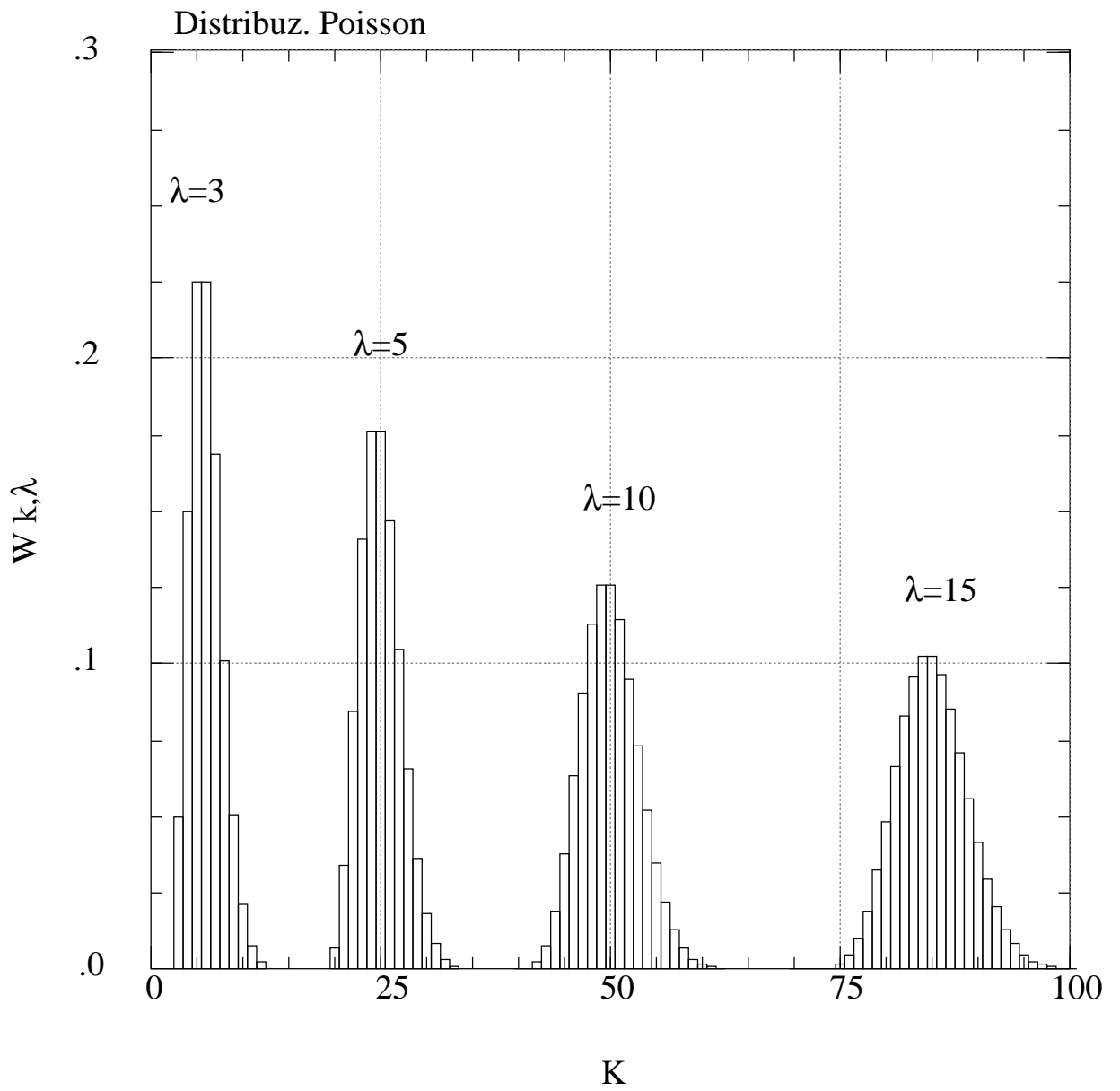
$$M_x(t) = E\{e^{tx}\} = \sum_{x=0}^{\infty} \frac{e^{tx} e^{-\lambda} \lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(e^t \lambda)^x}{x!} = e^{-\lambda} e^{\lambda e^t}$$

perchè è: $\sum_{x=0}^{\infty} \frac{e^{tx}}{x!} = e^{e^t}$.

moda

La moda è il valore di m per cui è massima la probabilità. Allora $P(m-1) < P(m)$ e $P(m+1) < P(m)$. Cioè

$$\frac{P(m-1)}{P(m)} < 1 \text{ e } \frac{P(m+1)}{P(m)} < 1$$



Distribuzione di Poisson per diversi valori di λ . I grafici sono traslati in ascissa per una migliore visibilità.

Ora è:

$$P(m) = \frac{\lambda^m}{m!} e^{-\lambda}; P(m-1) = \frac{\lambda^{m-1}}{(m-1)!} e^{-\lambda}; P(m+1) = \frac{\lambda^{m+1}}{(m+1)!} e^{-\lambda}$$

e

$$\frac{P(m-1)}{P(m)} = \frac{\lambda^{m-1}}{(m-1)!} e^{-\lambda} \cdot \frac{m!}{\lambda^m e^{-\lambda}} = \frac{m}{\lambda}$$
$$\frac{P(m+1)}{P(m)} = \frac{\lambda^{m+1}}{(m+1)!} e^{-\lambda} \cdot \frac{m!}{\lambda^m e^{-\lambda}} = \frac{\lambda}{m+1}$$

Per quanto detto sopra, dovrà essere contemporaneamente $\frac{m}{\lambda} < 1$ e $\frac{\lambda}{m+1} < 1$ cioè

$$m > \lambda - 1 \quad e \quad m < \lambda$$

da cui si deduce (λ , che è anche il valore di aspettazione di una variabile discreta, è un numero intero) che la distribuzione di Poisson è BIMODALE. Questa legge, per la sua caratteristica di esprimere la probabilità di un evento, la cui probabilità è piccola, nel caso di un gran numero di prove, è detta anche legge dei piccoli numeri o legge degli eventi rari.

ESERCIZIO:

In un nastro di stoffa prodotto industrialmente è presente in media 1 difetto ogni 1000 metri. Assumendo una distribuzione di Poisson per il numero di difetti in una data lunghezza, trovare la probabilità che

- a) un nastro di 2400 metri abbia al massimo 2 difetti
- b) un nastro di 1200 metri non abbia alcun difetto

Soluzione:

La probabilità media di un difetto è $1/1000=0.001$ difetti/metro.

- a) "un massimo" di 2 difetti significa 0,1 o 2 difetti. Allora

$$P(0) = \frac{(2400 \cdot 0.001)^0}{0!} e^{-2400 \cdot 0.001} = \frac{2.4^0}{0!} e^{-2.4} = 0.091$$

$$P(1) = \frac{2.4^1}{1!} e^{-2.4} = 0.2177$$

$$P(2) = \frac{2.4^2}{2!} e^{-2.4} = 0.2613$$

$$P(\text{al massimo } 2) = P(0) + P(1) + P(2) = 0.091 + 0.2177 + 0.2613 = 0.57$$

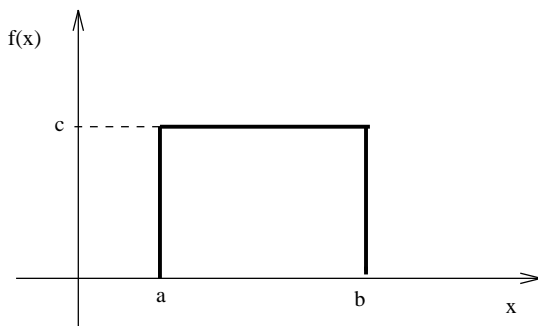
- b)

$$P(0) = \frac{1.2^0}{0!} e^{-1.2} = 0.30$$

DISTRIBUZIONE UNIFORME

Consideriamo ora il caso più semplice di distribuzione continua; quello in cui i valori della variabile aleatoria hanno la stessa densità di probabilità all'interno di un intervallo. Questa distribuzione è detta uniforme e la sua densità di probabilità è definita come:

$$\begin{aligned} f(x) &= c & a \leq x < b \\ f(x) &= 0 & x < a, x \geq b \end{aligned}$$



Dato che deve essere

$$\int_{-\infty}^{\infty} f(x)dx = c \int_a^b dx = c(b-a) = 1$$

ed essendo $f(x) = c$, sarà

$$f(x) = \frac{1}{b-a}$$

La funzione di distribuzione è

$$F(x) = \int_a^x \frac{dx}{b-a} = \frac{x-a}{b-a}$$

$$(a \leq x < b \quad P(a < X < x))$$

$$F(x) = 0 \quad \text{per } x < a$$

$$F(x) = 1 \quad \text{per } x \geq b$$

CARATTERISTICHE NUMERICHE

Il **valore di aspettazione** $E(x)$ è:

$$E(x) = \int_a^b \frac{x}{b-a} dx = \frac{1}{b-a} \int_a^b x dx = \frac{1}{b-a} \frac{b^2 - a^2}{2} = \frac{1}{b-a} \frac{(b-a)(b+a)}{2} = \frac{b+a}{2}$$

e la **varianza**

$$\sigma^2(x) = \int_a^b (x - \hat{x})^2 f(x) dx = \frac{1}{b-a} \int_a^b \left(x - \frac{b+a}{2}\right)^2 dx = \frac{(b-a)^2}{12}$$

Lo **scarto** è $\sigma(x) = \frac{b-a}{2\sqrt{3}}$.

Dato che il valore di aspettazione è a metà dell'intervallo della x , la distribuzione è simmetrica e lo **skewness** è nullo.

Il **kurtosis** deriva da

$$\mu_4 = \frac{1}{b-a} \int_a^b (x - \hat{x})^4 dx = \frac{(b-a)^4}{80}$$

ed è

$$kurt = \frac{\mu_4}{\sigma^4} - 3 = \frac{(b-a)^4}{80} \frac{2^4 3^2}{(b-a)^4} - 3 = \frac{9}{5} - 3 = -1.2$$

Funzione generatrice dei momenti La funzione generatrice dei momenti per la distribuzione uniforme è data da: $M_x(t) = E(e^{tx})$

$$M_x(t) = \int_{-\infty}^{+\infty} e^{tx} f(x) dx = \frac{1}{b-a} \int_a^b e^{tx} dx = \frac{e^{tb} - e^{ta}}{(b-a)t}$$

Questa formula, però, non è conveniente: meglio in questo caso il calcolo diretto dei momenti. La distribuzione uniforme viene utilizzata per generare eventi casuali indipendenti (pseudorandom), tramite un computer.

DISTRIBUZIONE NORMALE

La legge normale, o legge di Gauss, è una distribuzione che si applica ad una variabile aleatoria che sia la *risultante di un gran numero di cause indipendenti*, i cui effetti si sommano, delle quali nessuna è preponderante (valida anche nel caso in cui le leggi di distribuzione delle singole cause non siano normali). Condizioni di questo tipo si trovano in molte situazioni: errori di misura, diametro di pezzi fabbricati in serie, durata di un tragitto, fluttuazioni casuali di grandezze economiche (produzioni, vendita, ecc) intorno alla media. È dunque facile immaginare, dati anche i pochi vincoli cui è sottoposta, come la legge normale sia tra le più importanti leggi di probabilità. Essa è anzi la più importante distribuzione per *variabili continue*, così come la binomiale lo è per quelle discrete. Quando abbiamo una variabile **discreta** noi **contiamo**; quando abbiamo una variabile **continua** noi **misuriamo**. Con la misura siamo in grado di produrre tutti i valori possibili (almeno concettualmente) di una variabile aleatoria in un intervallo; infatti possiamo ad esempio misurare 67 o, con più accuratezza, 67.3, 67.28, 67.276, 67.2756, ...

una misura con bassa accuratezza ci permetterebbe forse di distinguere tra 67 e non 67 (binomiale). Rovesciando il ragionamento si può dire (e dimostrare) che la gaussiana è il caso limite della binomiale quando si hanno molti dati (> 30).

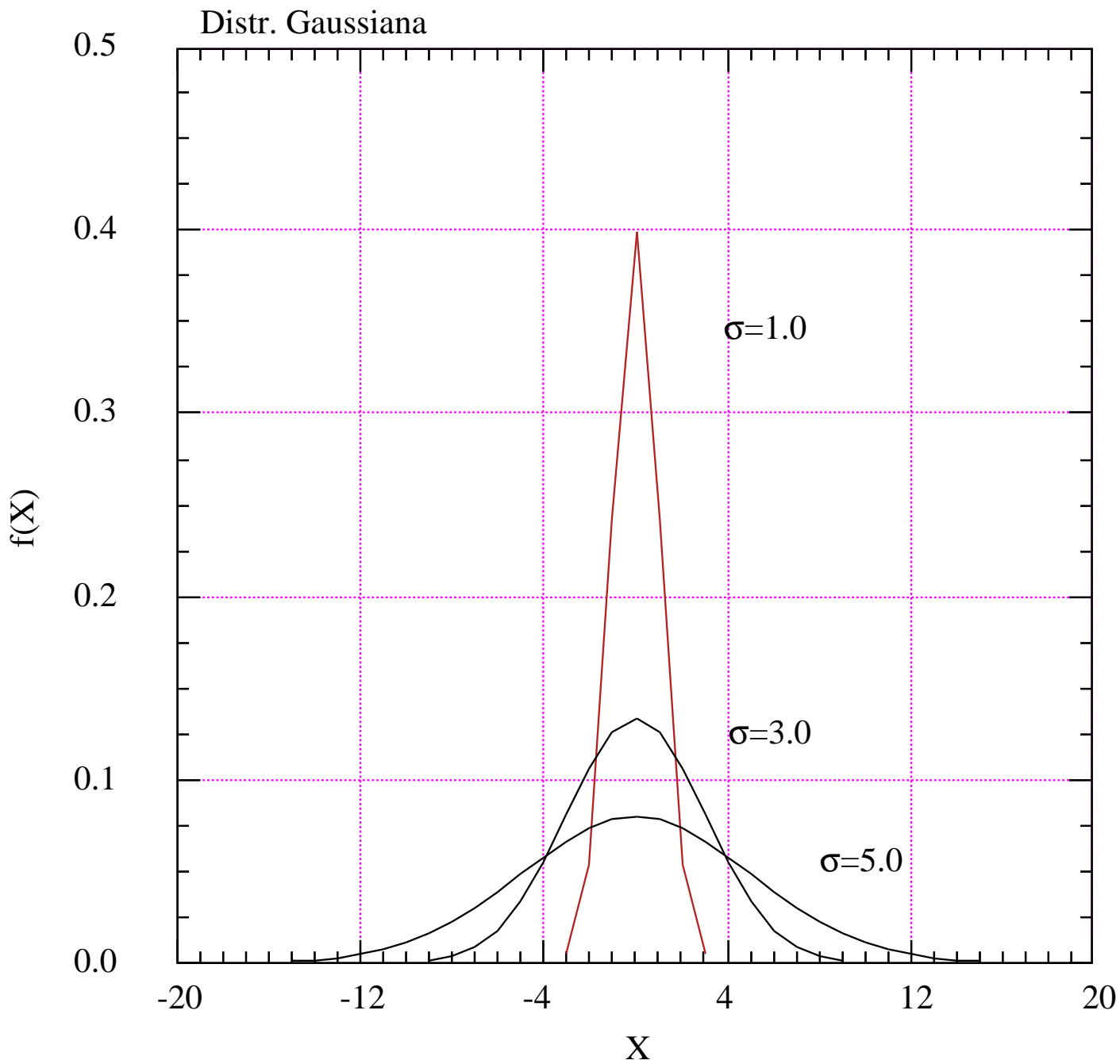
Noi non facciamo le funzioni caratteristiche, per cui non dimostreremo come si arriva ad una distribuzione normale. La sua espressione analitica è

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\hat{x})^2}{2\sigma^2}}$$

mentre la funzione di distribuzione è:

$$F(x) = P(X < x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(x-\hat{x})^2}{2\sigma^2}} dx$$

I parametri della distribuzione normale sono \hat{x} e σ^2 , valore di aspettazione e varianza dei dati. Al diminuire della dispersione σ , la distribuzione diviene più stretta.



La distribuzione normale al variare di sigma. Le tre curve sono centrate sul valore $\hat{x} = 0$

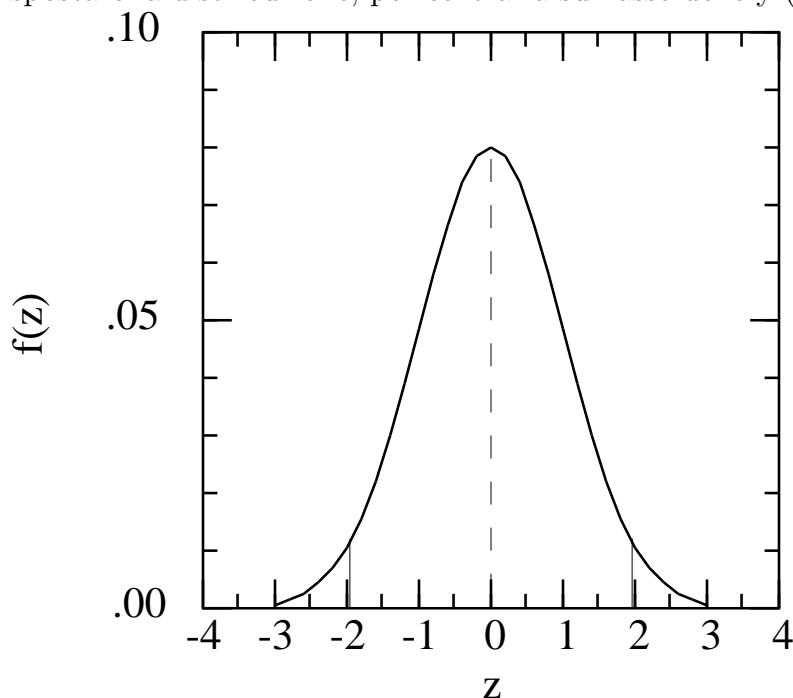
I calcoli successivi risultano più comprensibili se si passa alla funzione *centrata ridotta*, con il cambiamento di variabile:

$$z = \frac{x - \hat{x}}{\sigma}.$$

La densità di probabilità diventa allora:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}.$$

Il cambiamento di variabile ha il senso di esprimere la distanza tra un valore della variabile aleatoria e la media della popolazione in termini della deviazione standard σ , che in questo modo diventa "l'unità di misura" e quindi vale 1. Graficamente significa spostare la distribuzione, per centrarla sull'asse delle y (a $x=0$).



Diciamo quindi che *la distribuzione normale centrata ridotta ha media nulla e varianza unitaria*.

MOMENTI

I momenti centrali di ordine qualsiasi k sono, come al solito, dati da:

$$\mu_k = \int_{-\infty}^{\infty} (x - \hat{x})^k f(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x - \hat{x})^k e^{-\frac{(x-\hat{x})^2}{2\sigma^2}} dx$$

Ponendo $t = (x - \hat{x})/\sigma\sqrt{2}$ [$dt = dx/\sigma\sqrt{2}$], si ha:

$$\mu_k = \frac{\sigma\sqrt{2}}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} t^k (\sigma\sqrt{2})^k e^{-t^2} dt = \frac{(\sigma\sqrt{2})^k}{\sqrt{\pi}} \int_{-\infty}^{\infty} t^k e^{-t^2} dt \quad (A)$$

Ora è (v. ad esempio Mathematical Handbook, serie Shaum, pag. 98):

$$\int_0^{\infty} x^m e^{-ax^2} dx = \frac{\Gamma[(m+1)/2]}{2a^{(m+1)/2}}.$$

Per noi è $m=k$ e $a=1$ e quindi:

$$\mu_k = \frac{(\sigma\sqrt{2})^k}{\sqrt{\pi}} \Gamma[(k+1)/2].$$

Ora, μ_{k-2} è:

$$\mu_{k-2} = \frac{(\sigma\sqrt{2})^{k-2}}{\sqrt{\pi}} \Gamma[(k-1)/2]$$

e il loro rapporto:

$$\frac{\mu_k}{\mu_{k-2}} = 2\sigma^2 \frac{\Gamma[(k+1)/2]}{\Gamma[(k-1)/2]}$$

e quindi

$$\mu_k = 2\sigma^2 \frac{\Gamma[(k+1)/2]}{\Gamma[(k-1)/2]} \mu_{k-2}$$

che ci permette di calcolare i momenti, partendo dai momenti di ordine inferiore. Sapendo che $\mu_0 = 1$ e $\mu_1 = 0$ e ricordando (v. Brandt, App. D) che $\Gamma(x+1) = x\Gamma(x)$; $\Gamma(1) = 1$; $\Gamma(n+1) = n!$; $\Gamma(1/2) = \sqrt{\pi}$, si ricava ad esempio

$$\mu_2 = \sigma^2$$

$$\mu_3 = 0 \text{ (come tutti i momenti dispari)}$$

$$\mu_4 = 3\sigma^4$$

Lo SKEWNESS= $\mu_3/\sigma^3 = 0$, cioè la distribuzione normale è simmetrica; il KURTOSIS= $(\mu_4/\sigma^4) - 3 = (3\sigma^4/\sigma^4) - 3 = 0$, cioè il coefficiente di appiattimento è nullo, come sapevamo già, dato che riferiamo il kurtosis alla normale.

La probabilità che un certo valore della variabile aleatoria sia compreso tra $\pm\sigma$ è 0.682; tra $\pm 2\sigma$ è 0.954; tra $\pm 3\sigma$ è 0.998 (confrontare con le previsioni di Chebishev). Questo significa, ad esempio, che il 99.8% dei risultati di una prova sono compresi nell'intervallo $(-3\sigma \div +3\sigma)$.

Un'altra misura della larghezza della distribuzione normale è la *larghezza piena a mezza altezza* (*Full Width Half Maximum, contratto in FWHM*) pari a 2.35σ (v. esercizio successivo).

IMPORTANTE

Per convenzione, nell'ipotesi che gli errori di misura di una grandezza si distribuiscano normalmente, si adotta come incertezza (errore) della misura la deviazione standard σ della distribuzione degli errori. Quando viene data come risultato la quantità $(R \pm e_R)$, senza ulteriori precisazioni, si intende che e_R è la deviazione

standard della distribuzione normale che ha R come valore medio. La scrittura precedente significa anche che ci si aspetta, ripetendo l'esperimento, che il 68.2% dei risultati cada nell'intervallo $(R - e_R) \div (R + e_R)$.

Un commento: il modello degli errori esposto sopra, e tutti gli altri modelli possibili sono appunto MODELLI e non leggi di natura. Il fatto che la distribuzione degli errori segua la legge di Gauss è una cosa che va verificata di volta in volta perchè le cause degli errori sperimentali possono essere tali e tante da rendere non vere le ipotesi fatte: non sempre è facile assicurare la simmetria e l'indipendenza delle cause di errore.

ESERCIZIO:

Calcolare la FWHM di una distribuzione normale con media nulla e varianza 0.49 ($\sigma = 0.7$).

Soluzione:

Consideriamo metà curva: a metà altezza sarà $f(x) = \frac{f(0)}{2}$, ed essendo $f(0) = 1/\sqrt{2\pi}\sigma$, sarà $f(0)/2 = 1/2\sqrt{2\pi}\sigma$. Quindi $f(0)/2 = f(0)e^{-\frac{z^2}{2\sigma^2}}$. Allora

$$\frac{1}{2\sigma\sqrt{2\pi}} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{z^2}{2\sigma^2}}$$

e

$$\frac{1}{2} = e^{-\frac{z^2}{2\sigma^2}} \quad e^{-2\sigma^2 \ln \frac{1}{2}} = z^2$$

e quindi $z = 1.177\sigma$. Questo valore di z si riferisce all' HWHM (semi larghezza a mezza altezza) e quindi la FWHM è 2.354σ . Con i dati del problema è $FWHM = 2.354 \cdot 0.7 = 1.648$.

DISTRIBUZIONE χ^2

Se estraiamo un campione x_1, x_2, \dots, x_n , di dimensione n , da una distribuzione gaussiana di parametri a e σ^2 , è possibile dimostrare che la statistica (variabile aleatoria)

$$\chi^2 = \frac{(x_1 - a)^2 + (x_2 - a)^2 + (x_3 - a)^2 + \dots + (x_n - a)^2}{\sigma^2} = \sum_{i=1}^n \frac{(x_i - a)^2}{\sigma^2}$$

segue la funzione di distribuzione (*chi quadrato*)

$$F(\chi^2) = \frac{1}{\Gamma(\frac{\nu}{2})2^{\frac{\nu}{2}}} \int_0^{\chi^2} u^{\frac{\nu}{2}-1} e^{-\frac{u}{2}} du$$

con ν numero di gradi di libertà.

La corrispondente densità di probabilità è:

$$f(\chi^2) = \frac{1}{\Gamma(\frac{\nu}{2})2^{\frac{\nu}{2}}} (\chi^2)^{\frac{\nu}{2}-1} e^{-\frac{\chi^2}{2}}.$$

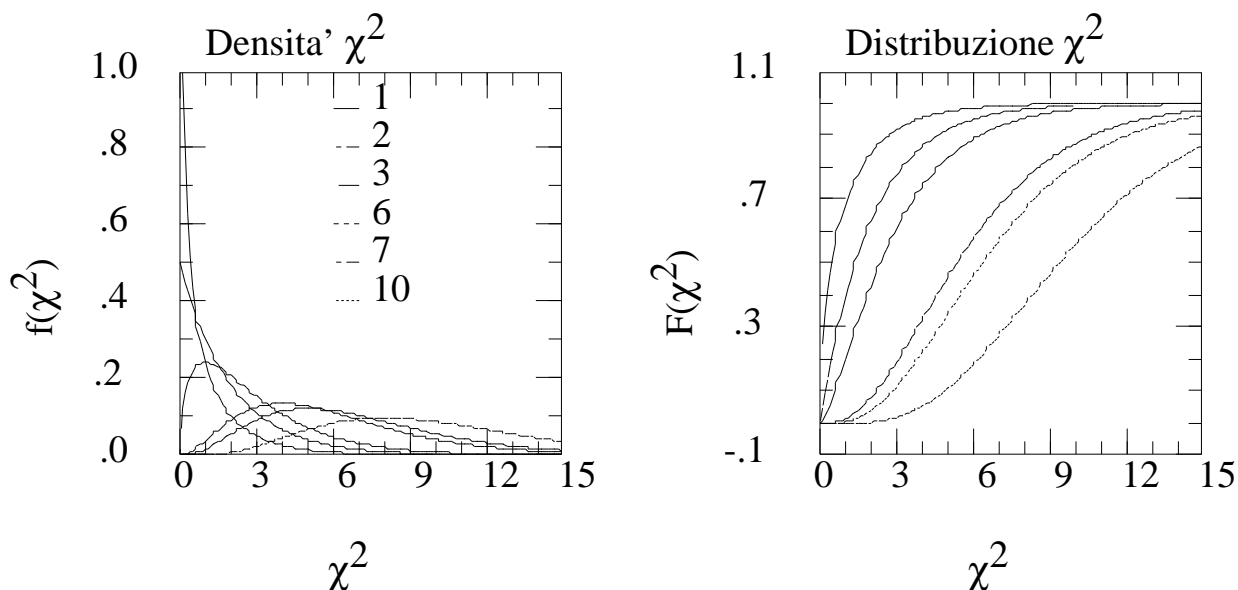
Si deve sottolineare che, avendo una distribuzione gaussiana, è possibile costruire una statistica (χ^2) che ha come unico parametro il numero dei gradi di libertà. Questo

equivale a dire che la "gaussianità" di un campione viene trasferita sulla distribuzione χ^2 , dipendendo da un solo parametro (e non più da due). Il valore di aspettazione della distribuzione χ^2 e la sua varianza sono:

$$E(\chi^2) = \nu \quad \sigma^2(\chi^2) = 2\nu$$

La $F(\chi^2)$ in funzione dei gradi di libertà, fornisce la probabilità che un campione con ν gradi di libertà sia compatibile con una distribuzione gaussiana.

Anche la χ^2 , per $\chi^2 \rightarrow \infty$, tende a diventare una gaussiana (teorema limite centrale). Dalla definizione di χ^2 si vede che se $\chi^2 = 0$ il dato sperimentale coincide con quello teorico. Maggiore è il valore di χ^2 , maggiore è la discrepanza dei dati dalla media. L'importanza del χ^2 si ritrova soprattutto nei test che confrontano una distribuzione osservata con una teorica.



Ricordiamo che la distribuzione χ^2 , con 1 grado di libertà, era stata ottenuta come applicazione del metodo della trasformazione di variabile.

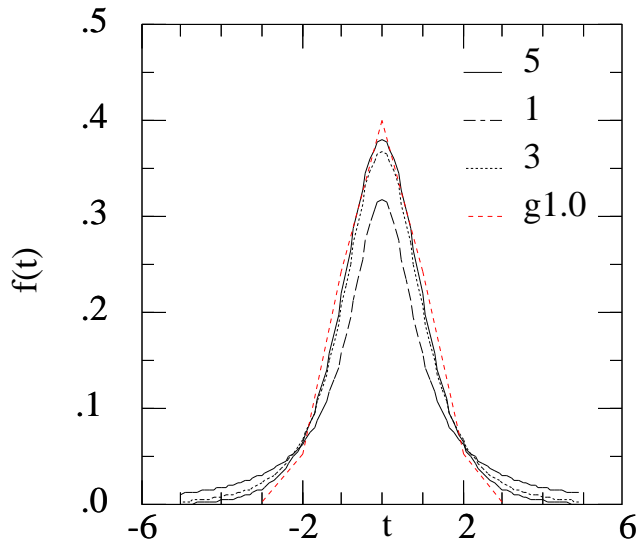
DISTRIBUZIONE di STUDENT

Una variabile aleatoria X segue una distribuzione di Student, che dipende solo dai gradi di libertà $\nu = N - 1$, se la sua densità è:

$$f(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi}\sqrt{\nu}} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

Si dirà allora che X segue una distribuzione di Student (o distribuzione t) e si indicherà $X \sim t(\nu)$.

La forma di questa densità è simile a quella della gaussiana standardizzata $N(0,1)$, ma è più bassa in centro e più alta nelle code.



I momenti sono:

$$E(x) = 0 \text{ e } var(x) = \frac{\nu}{\nu-2}$$

Uno dei motivi dell'importanza di questa distribuzione è il teorema seguente:

Siano X e Y due v.a. indipendenti, con $X \sim N(0,1)$ e $Y \sim \chi^2(\nu)$. La v.a.

$$T = \frac{X}{\sqrt{Y/\nu}}$$

segue la distribuzione di Student con ν gradi di libertà.

DISTRIBUZIONE di FISHER

Una v.a. continua X segue una distribuzione di Fisher se la sua densità di probabilità è:

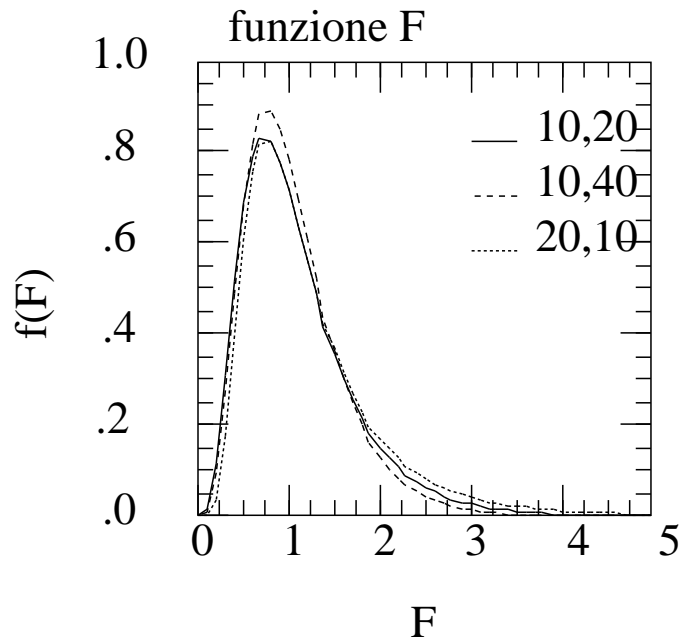
$$f(x) = \left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} \frac{\Gamma(\frac{\nu}{2})}{\Gamma(\frac{\nu_1}{2})\Gamma(\frac{\nu_2}{2})} x^{\frac{\nu_1}{2}-1} \left(1 + \frac{\nu_1}{\nu_2}x\right)^{-\frac{\nu}{2}},$$

con $\nu = \nu_1 + \nu_2$. Questa distribuzione dipende dai due parametri ν_1 e ν_2 .
il valore di aspettazione della v.a. di Fisher è:

$$E(x) = \frac{\nu_2}{\nu_2 - 2} \quad (\nu_2 > 2)$$

e la varianza;

$$Var(x) = \frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)} \quad (\nu_2 > 4)$$



Uno dei motivi dell'importanza di questa distribuzione è il teorema seguente:

Siano X_1 e X_2 due v.a. indipendenti, con $X_1 \sim \chi^2(\nu_1)$ e $X_2 \sim \chi^2(\nu_2)$. Si dimostra che la v.a.

$$X = \frac{X_1}{\nu_1} / \frac{X_2}{\nu_2}$$

segue la distribuzione di Fisher con ν_1 e ν_2 gradi di libertà.

Teorema Limite Centrale

Come abbiamo già visto, la distribuzione normale viene generata, sotto condizioni molto poco restrittive, per addizione di effetti numerosi ed indipendenti. Sia x_1, x_2, \dots, x_n una serie di variabili aleatorie che corrispondono agli effetti di cui sopra e che rispettano le seguenti condizioni:

- 1) le x_i sono indipendenti
- 2) i loro valori di aspettazione m_1, m_2, \dots, m_n e le loro varianze v_1, v_2, \dots, v_n esistono tutte
- 3) il rapporto tra la varianza di un elemento particolare della serie e la somma delle varianze $v_i / \sum v_i$ tende a zero per $n \rightarrow \infty$.

Chiamiamo $X = \sum_{i=1}^n x_i$ la somma di queste n variabili aleatorie. Sappiamo, dalle proprietà del valore di aspettazione che il valore di aspettazione della somma è uguale alla somma dei valori di aspettazione:

$$E\{X\} = E\left\{\sum_{i=1}^n x_i\right\} = \sum_{i=1}^n E\{x_i\} = m$$

con $m = m_1 + m_2 + \dots + m_n$.

La stessa cosa succede per le varianze:

$$V\{X\} = V\left\{\sum_{i=1}^n x_i\right\} = \sum_{i=1}^n V(x_i) = \sigma^2$$

con $\sigma^2 = V_1 + V_2 + \dots + V_n$.

Allora la condizione 3) può essere interpretata così: LA VARIABILITÀ DOVUTA AD OGNUNO DEGLI EFFETTI È PICCOLA RISPETTO ALLA VARIABILITÀ DOVUTA ALLA SOMMA DEGLI EFFETTI.

La variabile centrata ridotta z , costruita a partire dalla variabile X , è quindi

$$z = \frac{X - m}{\sigma} = \frac{\sum x_i - E\{\sum x_i\}}{\sqrt{\sum V\{x_i\}}}$$

Il teorema limite centrale afferma che questa variabile (z), quando $n \rightarrow \infty$, tende a seguire la legge normale, indipendentemente dalle leggi di probabilità seguite dalle singole variabili x_1, x_2, \dots, x_n .

Di conseguenza i fenomeni che possono essere considerati generati da un grande numero di cause elementari di fluttuazioni indipendenti, saranno facilmente rappresentabili con una legge normale.

Questo non sempre è vero:

- 1) il numero di cause indipendenti potrebbe essere troppo piccolo
- 2) i vari effetti potrebbero non essere additivi.

Le condizioni di applicazione del teorema limite centrale sono verificate in due casi particolarmente importanti:

- A) approssimazione della legge binomiale con la legge normale
- B) legge della media di un campione molto grande

A) Da binomiale a gaussiana

Se X è una variabile aleatoria binomiale $X = B(n, p)$ in cui $n \rightarrow \infty$ e p è intermedio tra 0 e 1, la binomiale tende alla normale di parametri $\sigma^2 = npq$ e $m = np$.

$$B(n, p) \rightarrow N(np, npq)$$

questo risultato si realizza per $n > 15, 20$.

La variabile X può essere considerata come la somma di n eventi elementari (variabili di Bernoulli) indipendenti $X = x_1 + x_2 + \dots + x_n$ che riuniscono le condizioni del teorema limite centrale

- 1) sono indipendenti
- 2) i valori di aspettazione e le varianze esistono

$$E\{x_i\} = p \qquad V\{x_i\} = pq$$

3) il rapporto tra varianza di un evento e somma delle varianze

$$\frac{V\{x_i\}}{\sum V\{x_i\}} = \frac{pq}{npq} = \frac{1}{n}$$

tende a zero per $n \rightarrow \infty$.

Quindi la variabile aleatoria X tende a seguire la legge normale di media $E\{X\} = \sum E\{x_i\} = np$ e varianza $V\{X\} = \sum V\{x_i\} = npq$.

B) Legge della media di un campione

Introduciamo il concetto di popolazione e di campione.

popolazione: l'insieme (eventualmente infinito) di tutti i possibili eventi. Segue la distribuzione teorica.

campione: un sottoinsieme di dimensione N della popolazione. Essenzialmente una serie finita di misure.

Se da una popolazione estraiamo, con rimessa, una serie di campioni (circa 30), risulterà che la loro media segue una legge normale di parametri m e σ^2/n , se m e σ^2 sono i momenti della popolazione.

$$\bar{X} \longrightarrow N(m, \sigma^2/n).$$

Infatti

$$\bar{X} = \frac{1}{n} \sum x_i = \frac{x_1}{n} + \frac{x_2}{n} + \dots + \frac{x_n}{n}$$

cioè la variabile aleatoria è la somma di n variabili aleatorie.

Le condizioni del teorema limite centrale sono;

- 1) le variabili sono indipendenti (estrazione con rimessa)
- 2) $E\{x_i\} = m$ (media della popolazione)
- $V\{x_i\} = \sigma^2$ (varianza della popolazione)
- 3)

$$\frac{V\{x_i\}}{\sum V\{x_i\}} = \frac{\sigma^2}{n\sigma^2} = \frac{1}{n} \rightarrow 0 \text{ per } n \rightarrow \infty.$$

Quindi \bar{X} (per $n \rightarrow \infty$) segue una legge normale di media= m e varianza= $V\{\bar{X}\} = V\{\frac{1}{n} \sum x_i\} = \frac{1}{n^2} \sum V\{x_i\} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$.

7. CAMPIONAMENTO

Nel capitolo precedente abbiamo illustrato alcune distribuzioni, ma non abbiamo definito come costruirle a partire dai dati sperimentali. Costruire una distribuzione vuol dire determinare i suoi parametri (λ per Poisson, σ e \hat{x} per Gauss, ecc.) che, per definizione, sono derivabili da una popolazione (cioè da un insieme, praticamente infinito, che copre tutti i casi possibili di un esperimento) e sono quindi sconosciuti. Non siamo quindi in grado di conoscere la distribuzione delle probabilità che dobbiamo approssimare con

una distribuzione di frequenza sperimentale. L'insieme delle misure che noi facciamo per poter ottenere questa approssimazione è detto campione. Se vengono fatte n misure, si dirà che il campione è di dimensione n .

Se la variabile aleatoria X segue la distribuzione $f(x)$, estrarre un campione (campionare la popolazione) vuol dire prendere un numero finito di valori di X , ed eventualmente ripetere l'esperimento più volte. Un singolo campione sarà un insieme $x_1, x_2, \dots, x_n = \mathbf{X}$ di valori della X e seguirà la densità di probabilità $g(\mathbf{X}) = g(x_1, x_2, \dots, x_n)$. Un campione deve essere aleatorio e deve soddisfare le condizioni

a) le x_i devono essere indipendenti, cioè $g(\mathbf{X}) = g(x_1)g(x_2) \dots g(x_n)$

b) le distribuzioni dei singoli elementi del campione devono essere uguali alla densità della popolazione: $g_1(x_1) = g_2(x_2) = \dots = g_n(x_n) = f(x)$

Avendo a disposizione un campione, si può definire n_x come il numero di elementi del campione per cui $X < x$ e quindi costruire la funzione

$$W_n(x) = \frac{n_x}{n}$$

che è una funzione di distribuzione empirica, detta distribuzione del campione. La $W_n(x)$ dovrebbe essere una approssimazione della distribuzione $F(x)$ della popolazione, alla quale tenderebbe per $n \rightarrow \infty$.

Sicuramente la distribuzione empirica conterrà uno o più parametri (come la distribuzione della popolazione) che saranno determinabili dai dati del campione (es. valore di aspettazione, varianza, momenti di ordine superiore).

Stima

Il problema di conoscere la funzione di distribuzione della popolazione a partire dalla funzione empirica, consiste allora nello stimare i parametri della popolazione, noti i parametri del campione. La stima è funzione delle variabili aleatorie x_1, x_2, \dots, x_n (gli elementi del campione), è cioè una statistica. Anche la stima è quindi una variabile aleatoria $S = S(x_1, x_2, \dots, x_n)$.

Precedentemente, definendo il valore di aspettazione, si è detto che questa grandezza può essere considerata il valore (non aleatorio) verso cui tende il risultato di un gran numero di esperimenti aleatori.

Evidentemente questo non è il caso dei campioni, dove la media è anch'essa aleatoria con una sua distribuzione di probabilità.

La stima, per poter essere considerata una valida rappresentazione della grandezza da stimare (e quindi per essere una **buona stima**) deve soddisfare alcune restrizioni:

- 1) al crescere del numero di prove (dimensione del campione), tende al parametro da stimare. In questo caso è detta **CORRETTA**.
- 2) la sostituzione del parametro con la sua stima non deve portare ad errori sistematici. **NON DISTORTA (UNBIASED)**.

3) la precisione della stima cresce all'aumentare del numero delle prove. In questo caso è detta **CONSISTENTE (o EFFICIENTE)**.

Chiamando S la stima del parametro λ , le condizioni precedenti si esprimono come:

- 1) $\lim_{n \rightarrow \infty} S_n = \lambda$ CORRETTA
- 2) $E\{S\} = \lambda$ NON DISTORTA
- 3) $\lim_{n \rightarrow \infty} \sigma(S) = 0$ CONSISTENTE (stima minima possibile)

Inoltre, date due stime "buone", una può essere più accettabile dell'altra, dato il contesto. Come confronto tra due stime S_1 e S_2 si usa il rapporto

$$\eta = \frac{\sigma^2(S_1)}{\sigma^2(S_2)}$$

e se $\eta < 1 \implies S_1$ migliore di S_2 .

Nei casi concreti non sempre si riescono a soddisfare le condizioni precedenti: può essere troppo complicato dare una stima efficiente e a volte si ricorre a stime con varianza maggiore. In alcuni casi un problema è risolvibile solo con stime distorte.

oooooooooooooooooooo

Classici esempi di stima sono la stima del valore di aspettazione e della varianza di una popolazione, a partire da un campione.

Sia data una variabile aleatoria X con valore di aspettazione \hat{x} e varianza σ^2 , INCOGNITI. Costruiamo, a partire da X , un campione di dimensione n (o n campioni di dimensione 1) x_1, x_2, \dots, x_n e formiamo la stima (corretta e non distorta) di \hat{x} (\bar{x}) e di σ^2 (s^2).

Come *valore di aspettazione* stimato prendiamo la *media aritmetica*

$$\bar{x} = S(\hat{x}) = \frac{\sum x_i}{n} \tag{1}$$

questa stima è corretta: per la legge dei grandi numeri $\lim_{n \rightarrow \infty} \bar{x} = \hat{x}$

non distorta: $E\{\bar{x}\} = \hat{x}$

La varianza di questa stima (v. teorema limite centrale) è

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \tag{2}$$

L'efficienza di una stima dipende dalla legge di distribuzione. Se X è distribuita normalmente, si può dimostrare che la (2) è minima, cioè \bar{x} è uno stimatore efficiente. Questo può non essere vero per altre leggi.

La *stima s^2 della varianza* della popolazione si può supporre data dalla definizione del momento centrale secondo:

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n}$$

Verifichiamo che questa sia una stima non distorta [$E(s^2) = \sigma^2$]:

$$\begin{aligned}
 E(s^2) &= \frac{1}{n} E \left\{ \sum_{i=1}^n (x_i - \bar{x})^2 \right\} \\
 &= \frac{1}{n} E \left\{ \sum_{i=1}^n (x_i - \hat{x})^2 - n(\bar{x} - \hat{x})^2 \right\} \\
 &= \frac{1}{n} \left\{ \sum_{i=1}^n E(x_i - \hat{x})^2 - nE(\bar{x} - \hat{x})^2 \right\} \\
 &= \frac{1}{n} \left\{ n\sigma^2 - n\frac{\sigma^2}{n} \right\} \\
 E(s^2) &= \left(1 - \frac{1}{n}\right) \sigma^2 = \left(\frac{n-1}{n}\right) \sigma^2
 \end{aligned} \tag{3}$$

da cui si vede che la stima è distorta. Per renderla non distorta si deve porre:

$$s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2.$$

Allora l'ultimo passaggio diventa

$$\frac{1}{n} \left\{ n\sigma^2 - n\frac{\sigma^2}{n} \right\} = \sigma^2 \left\{ \frac{n}{n-1} - \frac{1}{n-1} \right\} = \sigma^2$$

e quindi la stima della varianza (o varianza del campione) è:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2. \tag{4}$$

Il fattore $1/(n-1)$ è dovuto al fatto che gli n valori del campione sono già stati utilizzati per calcolare \bar{x} . Quindi, dato un campione, $n-1$ valori sono arbitrari, mentre l' n -esimo è fissato dal valore \bar{x} . In particolare è fissato da:

$$\sum_{i=1}^n (x_i - \bar{x}) = 0$$

(il momento centrale di ordine 1 è nullo).

Tutto questo si esprime sinteticamente dicendo che il numero di gradi di libertà della somma di quadrati è $n-1$. La somma di quadrati $\sum_{i=1}^n (x_i - \bar{x})^2$ divisa per i gradi di libertà (ν) [**eq. (4)**] si chiama *scarto quadratico medio* e la sua radice quadrata *scarto rms (root mean square)*.

In generale il numero dei gradi di libertà è dato da (numero di dati - numero dei vincoli [parametri])

$$\nu = n - p$$

Nota: Passaggi relativi all'equazione (3) precedente.

$$\sum (x_i - \bar{x})^2 = \sum [(x_i - \hat{x}) - (\bar{x} - \hat{x})]^2 = \sum (x_i - \hat{x})^2 - 2(\bar{x} - \hat{x}) \sum (x_i - \hat{x}) + n(\bar{x} - \hat{x})^2$$

Ora, la sommatoria a secondo membro è

$$\sum (x_i - \hat{x}) = \sum x_i - n\hat{x}$$

da cui, moltiplicando e dividendo per n ,

$$n \left(\frac{\sum x_i}{n} - \frac{n\hat{x}}{n} \right) = n(\bar{x} - \hat{x})$$

Quindi:

$$\sum (x_i - \bar{x})^2 = \sum (x_i - \hat{x})^2 - 2n(\bar{x} - \hat{x})^2 + n(\bar{x} - \hat{x})^2 = \sum (x_i - \hat{x})^2 - n(\bar{x} - \hat{x})^2$$

cioè il risultato precedente

Dalla (2), si può calcolare la **varianza della media**

$$var(\bar{x}) = \frac{s^2}{n} = \frac{\sum_i^n (x_i - \bar{x})^2}{n(n-1)} \quad (5)$$

e la **deviazione standard della media** come

$$s(\bar{x}) = \sqrt{\frac{\sum_i^n (x_i - \bar{x})^2}{n(n-1)}} \quad (6)$$

nota: nel caso di n osservazioni indipendenti, $X_{i,k}$ della grandezza X_i , ottenute nelle stesse condizioni di misura, la stima x_i si ottiene usualmente dalla media del campione

$$\bar{x}_i = \bar{X}_i = \frac{\sum x_i}{n}.$$

L'*incertezza* associata con x_i è la deviazione standard stimata della media

$$s(\bar{x}) = \sqrt{\frac{\sum_i^n (x_i - \bar{x})^2}{n(n-1)}}$$

Stima dei parametri della regressione lineare

Applichiamo la stima e la verifica dei criteri per una buona stima al caso della regressione lineare.

Supponiamo di osservare, per i da 1 a n , le risposte y_i in corrispondenza dei valori immessi x_i e di voler usare queste risposte per stimare i parametri α e β della regressione lineare $y_i = \alpha + \beta x_i$ tramite i **minimi quadrati**. Se chiamiamo A e B le stime di α e β , la stima della risposta é $y_i = A + Bx_i$ e la quantità $R_i = y_i - A - Bx_i$ é la differenza tra osservato e calcolato (**residuo**). La somma dei quadrati dei residui é allora:

$$SS = \sum_{i=1}^n (y_i - A - Bx_i)^2.$$

che rendiamo minima derivando rispetto ad A e B e uguagliando a zero.

$$\frac{\partial SS}{\partial A} = -2 \sum (y_i - A - Bx_i) = 0$$

(1)

$$\frac{\partial SS}{\partial B} = -2 \sum x_i (y_i - A - Bx_i) = 0$$

da cui si ricava il sistema di equazioni:

$$\sum y_i = nA + B \sum x_i$$

(2)

$$\sum x_i y_i = A \sum x_i - B \sum x_i^2$$

Ponendo $\bar{y} = \frac{1}{n} \sum y_i$ e $\bar{x} = \frac{1}{n} \sum x_i$, la prima delle (2) diventa $A = \bar{y} - B\bar{x}$ che, sostituita nella seconda delle (2), dá:

$$\sum x_i y_i = (\bar{y} - B\bar{x}) \cdot n\bar{x} + B \sum x_i^2$$

o

$$B(\sum x_i^2 - n\bar{x}^2) = \sum x_i y_i - n\bar{x} \cdot \bar{y}$$

da cui

$$B = \frac{\sum x_i y_i - n\bar{x} \cdot \bar{y}}{\sum x_i^2 - n\bar{x}^2}.$$

Quindi gli stimatori dei minimi quadrati di α e β sono

$$A = \bar{y} - B\bar{x}$$

e

$$B = \frac{\sum x_i y_i - n\bar{x} \cdot \bar{y}}{\sum x_i^2 - n\bar{x}^2}.$$

$$(n\bar{x} \cdot \bar{y} = n\bar{x} \frac{\sum y_i}{n} = \bar{x} \sum y_i)$$

Vediamo ora se A e B sono stimatori **non distorti** [E(S)=λ] di α e β. B può essere riscritto come

$$B = \frac{\sum (x_i - \bar{x}) y_i}{\sum x_i^2 - n\bar{x}^2}$$

da cui si vede che B é una combinazione lineare delle v.a. normali y_i e che quindi ha distribuzione normale con parametri:

$$\begin{aligned} E(B) &= \frac{\sum (x_i - \bar{x}) E(y_i)}{\sum x_i^2 - n\bar{x}^2} = \frac{\sum (x_i - \bar{x}) (\alpha + \beta x_i)}{\sum x_i^2 - n\bar{x}^2} = \frac{\alpha \sum (x_i - \bar{x}) + \beta \sum x_i (x_i - \bar{x})}{\sum x_i^2 - n\bar{x}^2} = \\ &= \frac{\beta \sum x_i^2 - n\bar{x}^2}{\sum x_i^2 - n\bar{x}^2} = \beta \dots c.v.d. \end{aligned}$$

($\sum (x_i - \bar{x}) = 0$ perché é il momento centrale di ordine 1)

Quindi B é uno stimatore non distorto di β. Adesso calcoliamo la varianza di B:

$$\begin{aligned} Var(B) &= Var \frac{\sum (x_i - \bar{x}) y_i}{\sum x_i^2 - n\bar{x}^2} = \frac{Var(\sum (x_i - \bar{x}) y_i)}{(\sum x_i^2 - n\bar{x}^2)^2} = \\ &= \frac{\sum (x_i - \bar{x})^2 Var(y_i)}{(\sum x_i^2 - n\bar{x}^2)^2} = \frac{\sigma_y^2 \sum (x_i - \bar{x})^2}{(\sum x_i^2 - n\bar{x}^2)^2} = \frac{\sigma_y^2}{\sum x_i^2 - n\bar{x}^2} \end{aligned}$$

perché é: $\sum (x_i - \bar{x})^2 = \sum x_i^2 - n\bar{x}^2$

Facciamo la stessa cosa per A:

$$A = \frac{1}{n} \sum y_i - B\bar{x}$$

Anche A segue una normale di parametri:

$$E(A) = \frac{1}{n} \sum E(y_i) - \bar{x} E(B) = \frac{1}{n} \sum (\alpha + \beta x_i) - \bar{x} \beta = \alpha + \beta \bar{x} - \bar{x} \beta = \alpha \dots c.v.d.$$

A é uno stimatore non distorto di α. La varianza di A vale:

$$Var(A) = \frac{\sigma_x^2 \sum x_i^2}{n(\sum x_i^2 - n\bar{x}^2)}$$

Abbiamo visto in precedenza che un campione x_1, x_2, \dots, x_n , di dimensione n , estratto da una distribuzione gaussiana di parametri a e σ^2 , dal quale si costruisca la statistica (variabile aleatoria)

$$\chi^2 = \sum_{i=1}^n \frac{(x_i - a)^2}{\sigma^2}$$

segue la funzione di distribuzione $\chi^2(\nu)$ [*chi quadrato*] con ν numero di gradi di libertà.

- Detta s^2 la stima di σ^2 , si può dimostrare che la quantità (la statistica)

$$U = \frac{n-1}{\sigma^2} s^2$$

segue una distribuzione χ^2 con $n-1$ gradi di libertà.

8. IL METODO DELLA MASSIMA VEROSIMIGLIANZA

Nel capitolo precedente abbiamo visto cosa significhi fare un campionamento e abbiamo definito il concetto di stima. Abbiamo anche visto come stimare il valore di aspettazione e la varianza, partendo da campioni tratti da popolazioni infinite. Abbiamo costruito una variabile (χ^2) a partire da una popolazione gaussiana, utile perchè legata solo ai gradi di libertà.

Adesso parleremo del problema generale della *stima dei parametri di una distribuzione*. Sia $\underline{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_p)$ un vettore le cui componenti siano i p parametri da stimare, che caratterizzano una distribuzione (*o una legge fisica*); $\mathbf{X} = (x_1, x_2, \dots, x_n)$ le variabili aleatorie considerate e $f(\mathbf{X}, \underline{\lambda})$ la densità di probabilità, funzione anche dei parametri.

ESEMPIO: se abbiamo a che fare con una distribuzione gaussiana, $\underline{\lambda} = (\lambda_1, \lambda_2) = a, \sigma$. $\mathbf{X} = (x_1, \dots, x_n)$ sono le misure e

$$f(\mathbf{X}, \underline{\lambda}) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - a)^2}{2\sigma^2}}$$

Estraiamo ora un campione da \mathbf{X} , che chiameremo \mathbf{X}_{sper} (oppure, ed è la stessa cosa, facciamo una serie di misure di \mathbf{X}). Supponiamo che il campione abbia **dimensione 1**.

- Come possiamo determinare i parametri $\underline{\lambda}$ in modo che l'ipotesi fatta (cioè la scelta della distribuzione) sia in accordo con il campione ?

Scegliamo i valori di $\underline{\lambda}$ in modo che la funzione $f(\mathbf{X}_{sper}, \underline{\lambda}_{stim})$, calcolata nel punto x_{sper} abbia massima probabilità. Cioè:

$$dP_{sper} = f(\mathbf{X}_{sper}, \underline{\lambda}_{stim}) d\mathbf{X} = \text{massima}.$$

(Il significato di "sper" e "stim" è : "sperimentale" e "stimato").

Se il campione ha **dimensione 2** ($x_{1,sper}, x_{2,sper}$), come sceglieremo i parametri? Esattamente nello stesso modo. A priori non c'è alcun modo di stabilire quale dei due risultati sia il più probabile e il fatto che proprio quei valori (e non altri) siano il risultato, ci autorizza a pensare che entrambi siano ugualmente probabili. Cercheremo quindi i valori di λ tali che la probabilità di avvenire sia massima *contemporaneamente per i due valori*. Quindi $dP_1 \cdot dP_2 = \text{massima}$.

In generale, per un campione di dimensione N sarà:

$$dP = \prod_{j=1}^N f(\mathbf{X}_j, \lambda_{stim}) d\mathbf{X} \quad \text{massima.}$$

La quantità dP , calcolata nei punti sperimentali, non è una vera e propria probabilità. È chiamata probabilità a posteriori e rappresenta la probabilità di accadere di un evento, calcolata dopo che l'evento si è presentato (... ma in questo caso l'evento ha probabilità 1). *La probabilità a posteriori non va mai confusa con la probabilità.*

La funzione

$$L = \prod_{j=1}^N f(\mathbf{X}_j, \lambda)$$

è detta funzione di verosimiglianza.

Il principio di massima verosimiglianza stabilisce che la *migliore stima dei parametri è quella che rende massima la funzione di verosimiglianza*.

Ma più di un insieme di parametri può rendere massima la L (cioè la L può avere più massimi, eventualmente della stessa altezza). Per confrontare due set di parametri si usa il quoziente di verosimiglianza

$$Q = \frac{\prod_{j=1}^N f(\mathbf{X}_j, \lambda_1)}{\prod_{j=1}^N f(\mathbf{X}_j, \lambda_2)}$$

calcolato il quale, diremo che il set di parametri λ_1 è Q volte più probabile del set λ_2 .

ESEMPIO

Dal lancio di una moneta asimmetrica si deve decidere a quale delle due classi appartiene:

	A	B
testa	1/3	2/3
croce	2/3	1/3

Da un campione di 5 lanci si ottiene 1 volta testa e 4 volte croce. Allora

$$L_A = \frac{1}{3} \left(\frac{2}{3}\right)^4 ; \quad L_B = \frac{2}{3} \left(\frac{1}{3}\right)^4$$

$$Q = \frac{\frac{1}{3} \left(\frac{2}{3}\right)^4}{\frac{2}{3} \left(\frac{1}{3}\right)^4} = \frac{\left(\frac{2}{3}\right)^3}{\left(\frac{1}{3}\right)^3} = 8.$$

È molto probabile che la moneta appartenga alla classe A (8 volte di più che appartenere alla classe B).

Per rendere massima la funzione di verosimiglianza si calcolano le sue derivate rispetto ai parametri e le si uguagliano a zero. Le derivate di una funzione prodotto di funzioni, come L , possono essere molto complicate. Si preferisce calcolare la funzione logaritmica di verosimiglianza l

$$l = \ln L = \sum_{j=1}^N \ln f(\mathbf{X}_{j, \text{super}}, \lambda_{\text{stim}}).$$

Questa funzione (a volte viene chiamata funzione di verosimiglianza, come L) ha gli estremi in corrispondenza degli estremi della L , e quindi il problema di massimizzare la L si esprime con

$$\frac{\partial l}{\partial \lambda_i} = 0 \quad i = 1, 2, \dots, p$$

ESEMPIO (**Importante**)

Consideriamo un caso di particolare importanza per la fisica: quello di una **serie di misure con precisione diversa**. Ogni misura corrisponde quindi ad aver campionato una distribuzione gaussiana con valore medio \underline{a} e varianza σ_j^2 . Il valore medio a , comune a tutte le misure, è il parametro da determinare.

Soluzione:

La probabilità a posteriori è

$$f(\mathbf{X}_{\text{super}}, \lambda) d\mathbf{X} = \frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{(x_j - a)^2}{2\sigma_j^2}} dx$$

Se le misure sono N :

$$L = \prod_{j=1}^N \frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{(x_j - a)^2}{2\sigma_j^2}}$$

$$l = \ln L = -\frac{1}{2} \sum_{j=1}^N \frac{(x_j - a)^2}{\sigma_j^2} + \text{cost}$$

e

$$\frac{dl}{da} = \sum_{j=1}^N \frac{(x_j - a)}{\sigma_j^2} = 0,$$

da cui

$$\sum_{j=1}^N \frac{x_j}{\sigma_j^2} - a \sum_{j=1}^N \frac{1}{\sigma_j^2} = 0$$

e

$$a = \frac{\sum_{j=1}^N \frac{x_j}{\sigma_j^2}}{\sum_{j=1}^N \frac{1}{\sigma_j^2}}$$

cioè la media pesata (essendo il peso $= \frac{1}{\sigma^2}$).

È possibile dimostrare (ad esempio applicando la propagazione degli errori all'equazione precedente) che la varianza della stima, $\sigma^2(a)$, vale:

$$\sigma^2(a) = \left(\sum_{j=1}^N \frac{1}{\sigma_j^2} \right)^{-1}$$

Bisogna notare che le due equazioni precedenti, nel caso in cui le misure x_j abbiano tutte lo stesso errore $\sigma(x)$, si riducono a: $a = \bar{x}$ e $\sigma^2(a) = \sigma^2(x)/N$. In particolare, ricaviamo ancora che *la varianza della media di un campione è uguale alla varianza della popolazione diviso la dimensione del campione.*

Tralasciamo la trattazione teorica che riguarda le proprietà degli stimatori derivati dalla massima verosimiglianza (stime **corrette**, **non distorte**, e a **minima varianza**).

Ricordiamo ancora che in molti casi pratici si impone un compromesso tra **bias** e **minima varianza**.

FIT CON LA MASSIMA VEROSIMIGLIANZA

Nel caso in cui il parametro da determinare sia funzione di altri parametri, il problema non cambia e la massimizzazione viene fatta rispetto ai nuovi parametri. Quindi, una volta fissata la densità di probabilità, trovare un parametro di una legge fisica (o di una relazione funzionale) vuol dire trovare un parametro della popolazione parente a partire dalla sua campionatura (di dimensione 1) che ha come risultato il valore sperimentale. Consideriamo come esempio pratico il caso in cui i campioni (i dati sperimentali) siano estratti da una popolazione normale e leghino due grandezze tramite una relazione lineare (supposta vera). *Quindi tutti i dati sono campionature di $f(x)$, una popolazione normale il cui valore medio (val. di aspettazione) è legato ad essi dalla relazione $y = a + bx$, con a e b parametri da determinare.*

La probabilità a posteriori è, in questo caso:

$$f(\mathbf{X}_{sper}, \lambda) d\mathbf{X} = \frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{(y_j - (a + bx_j))^2}{2\sigma_j^2}} dy$$

Se le misure sono N, la funzione di verosimiglianza è:

$$L = \prod_{j=1}^N \frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{(y_j - (a + bx_j))^2}{2\sigma_j^2}}$$

e quella logaritmica:

$$l = \ln L = -\frac{1}{2} \sum_{j=1}^N \frac{(y_j - (a + bx_j))^2}{\sigma_j^2} + \text{cost}$$

Ora, la massimizzazione della funzione logaritmica equivale a minimizzare l'esponente di e . Quindi

$$\sum_{j=1}^N \frac{[y_j - (a + bx_j)]^2}{\sigma_j^2} \text{ minimo.}$$

σ_j^2 è una costante, quindi bisogna rendere minima la quantità $\sum [y_j - (a + bx_j)]^2$. Cioè bisogna *rendere minima la somma dei quadrati dei residui* (avendo definito *residuo* la differenza tra valore sperimentale e valore "teorico" dedotto dalla relazione assunta come vera). Questo è il **principio dei minimi quadrati**. I minimi quadrati possono quindi essere considerati un caso particolare della massima verosimiglianza, quando la distribuzione parente è normale.

- Se le misure hanno precisione diversa (e quindi le σ_j sono diverse), detti $p_j = \frac{1}{\sigma_j^2}$ i pesi, si dovrà minimizzare

$$\sum_{j=1}^N p_j [y_j - (a + bx_j)]^2.$$

per ottenere i *minimi quadrati pesati*.

ESEMPIO:

Avendo fissato i parametri "tecnici" di un cannone, come l'alzo, il tipo di proiettile, ecc., si osservi il risultato di N colpi. Si supponga che i diversi valori osservati x_1, \dots, x_N siano distribuiti normalmente attorno ad una certa gittata media g , con deviazione standard σ .

Si vuole stimare dai dati sperimentali la migliore approssimazione di g e σ , che chiameremo \bar{g} e $\bar{\sigma}$.

Soluzione:

La funzione di verosimiglianza è:

$$L = \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - g)^2}{2\sigma^2}}$$

e quella logaritmica:

$$l = \ln L = -\frac{1}{2} \sum_{j=1}^N \frac{(x_i - g)^2}{\sigma^2} - N \ln \sigma - \frac{N}{2} \ln 2\pi$$

Le derivate rispetto ai parametri g e σ sono:

$$\frac{\partial l}{\partial g} = \sum_{i=1}^N \frac{x_i - g}{\sigma^2}$$

$$\frac{\partial l}{\partial \sigma} = \frac{1}{\sigma^3} \sum_{i=1}^N (x_i - g)^2 - \frac{N}{\sigma}$$

uguagliando a zero le due espressioni, si ottiene

$$\sum \frac{x_i - g}{\sigma^2} = 0 \implies \sum x_i - Ng = 0 \implies g = \frac{\sum x_i}{N}$$

cioè la media dei valori, e

$$\frac{1}{\sigma^3} \sum (x_i - g)^2 - \frac{N}{\sigma} = 0 \implies \frac{\sum (x_i - g)^2}{\sigma^2} = N$$

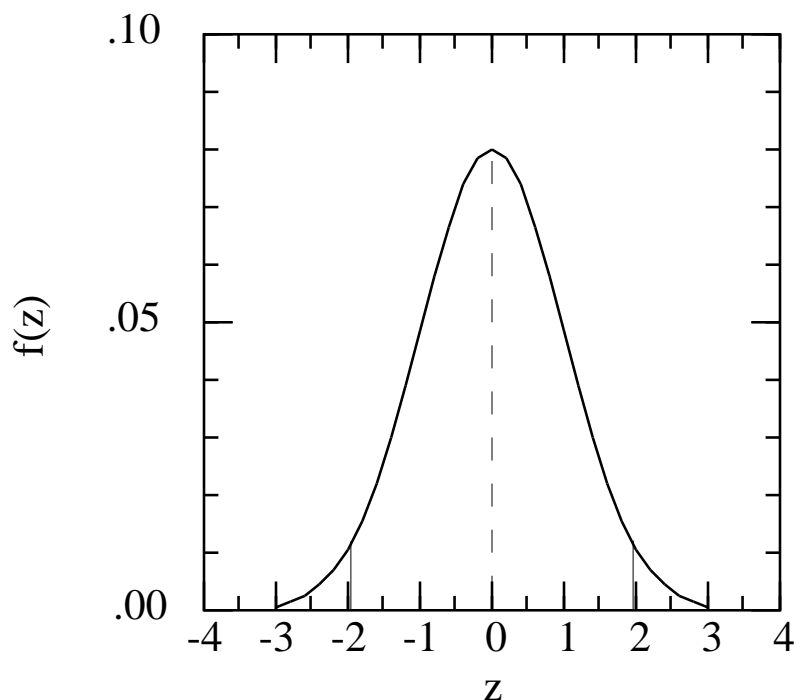
e quindi

$$\sigma = \sqrt{\frac{\sum (x_i - g)^2}{N}}.$$

STIME PER INTERVALLO (da non fare)

Se un parametro viene stimato fornendo un solo numero, si dice che è stata fatta una stima puntuale del parametro. Se invece la stima fornisce **due valori**, entro i quali si può supporre si trovi il valore "vero" del parametro, si parla di stima per intervallo. Ad esempio, se calcoliamo la media \bar{x} e la deviazione standard $s_{\bar{x}}$ di un campione x_1, x_2, \dots, x_n , potremmo dire che la stima della media della popolazione \hat{x} sarà compresa tra $\bar{x} - s_{\bar{x}}$ e $\bar{x} + s_{\bar{x}}$, con una probabilità del 68.2 % (se la distribuzione della x è normale). In questo caso avremmo generato una stima per intervallo della media della popolazione, stima dotata di un'affidabilità del 68.2 %.

In generale, la dimensione dell'intervallo entro cui stimiamo la presenza del parametro da trovare è data da $\bar{x} \pm z \cdot s_{\bar{x}}$ (nel caso di distribuzione normale), in cui z è l'ascissa tale che l'integrale da z all' ∞ della gaussiana sia uguale a $(1 - \text{il livello di affidabilità})$. Questa quantità, una probabilità, viene chiamata α . L'affidabilità, $(1 - \alpha)$, si chiama livello di confidenza. Diremo che siamo fiduciosi di trovare il parametro, nell'intervallo definito, con livello di confidenza $(1 - \alpha)$. Questo significa che, *a lungo andare*, in una percentuale $100(1 - \alpha)\%$ di campioni, \hat{x} risulterà compreso nell' intervallo di confidenza calcolato sopra.



Nella maggioranza dei casi concreti si avrà un solo campione e quindi un solo intervallo di confidenza che potrà essere uno dei $100(1 - \alpha)/100$ che includono il valore "vero" di \hat{x} , oppure uno dei $100\alpha/100$ che non lo includono. Lo stesso ragionamento seguito finora si può usare per determinare la stima per intervallo (o l'*intervallo di confidenza della stima*) di altri parametri delle distribuzioni.

ESEMPIO 1 - Stima del valore di aspettazione \hat{x} di una variabile aleatoria (v.a.) normale con varianza σ^2 nota.

La stima viene fatta a partire da un campione x_1, x_2, \dots, x_n osservato, con media \bar{x} e varianza (del campione) $\frac{\sigma^2}{n}$. In questo caso, come visto sopra, l'intervallo di confidenza per \hat{x} al livello di confidenza $(1 - \alpha)$ è:

$$\left[\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right]$$

ESEMPIO 2 - Stima del valore di aspettazione \hat{x} di una v.a. normale con varianza σ^2 ignota.

In questo caso usiamo la statistica $t = \frac{\bar{x} - \hat{x}}{s/\sqrt{n}}$ ($s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$, stima corretta della varianza) che segue la distribuzione di Student.

L'intervallo di confidenza per \hat{x} , al livello di confidenza $(1 - \alpha)$ è

$$\left[\bar{x} - t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}, \bar{x} + t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \right]$$

Nota: È $t = \frac{\bar{x} - \hat{x}}{s/\sqrt{n}} \longrightarrow \hat{x} = \bar{x} \pm \frac{st}{\sqrt{n}}$

ESEMPIO 3 - Stima della varianza σ^2 di una v.a. normale con valore di aspettazione \hat{x} ignoto.

In questo caso usiamo la statistica $Y = \frac{\nu s^2}{\sigma^2}$, con ν numero dei gradi di libertà, che segue la distribuzione $\chi^2 \left[\sigma^2 = \frac{\nu s^2}{y} \right]$.

L'intervallo di confidenza per σ^2 , al livello di confidenza $(1 - \alpha)$, è allora:

$$\left[\frac{\nu s^2}{y_{2,\alpha/2}}, \frac{\nu s^2}{y_{1,\alpha/2}} \right]$$

in cui $y_{\nu,\alpha/2}$ corrisponde a $z_{\alpha/2}$, tenendo presente che la distribuzione χ^2 non è simmetrica e che quindi è necessario calcolare sia y_1 che y_2 .

Determinazione della dimensione del campione

Riprendendo l'Esempio 1, è facile notare che l'ampiezza A dell'intervallo di confidenza, fissato $(1 - \alpha)$, è

$$A = \frac{2z_{\alpha/2}\sigma}{\sqrt{n}},$$

da cui, fissata l'ampiezza dell'intervallo e nota la varianza, si può ricavare la numerosità (dimensione) del campione:

$$n = \left[\frac{2z_{\alpha/2}\sigma}{A} \right]^2.$$

Ad esempio, se $\sigma^2 = 4$ e se si vuole che al livello $(1 - \alpha) = 95\%$ l'ampiezza dell'intervallo sia uguale a 2, si deve estrarre un campione di dimensione

$$n = \left[\frac{2 \cdot 1.96 \cdot 2}{2} \right]^2 \simeq 16.$$

Se la varianza non è nota, per poter avere una approssimazione di \underline{n} bisogna estrarre un campione ridotto (si fa in questo caso un'indagine "pilota") da cui ottenere una stima di $\sigma^2 = s^2$. Allora \underline{n} si trova da:

$$n = \left[\frac{2t_{\alpha/2}s}{A} \right]^2.$$

9. TEST DELLE IPOTESI STATISTICHE

Un'ipotesi è una affermazione sullo stato della natura. In molti campi, anche della vita di tutti i giorni, vengono prese decisioni sulla base di ipotesi, ipotesi che hanno bisogno di essere considerate vere (accettate) o false (respinte). Per questo fine le ipotesi hanno bisogno di dati che le confermino o meno. Questo processo è detto "test delle ipotesi".

- Un'ipotesi statistica può essere molto precisa e dettagliata, asserendo che lo stato della natura (in un particolare contesto) è completamente definito da un modello di probabilità. Si parla allora di **ipotesi semplice**.
- Oppure può essere descritta in modo più generale, facendo riferimento a condizioni che sono valide per più di un modello di probabilità o per più di una situazione nello stesso modello. Questa ipotesi è detta **composta**. Si assume che un'ipotesi composta sia formata da più ipotesi semplici.

ESEMPIO: L'affermazione che X è normalmente distribuita è un'ipotesi composta, perchè la proprietà di normalità non definisce completamente la distribuzione di X ; include ad esempio l'ipotesi semplice che $\hat{x} = 2$ e $\sigma^2 = 8$. L'affermazione che la variabile Y ha media 2 e varianza 8 è composta perchè questi due momenti da soli non definiscono una distribuzione.

"Testare" un'ipotesi vuol dire condurre un esperimento casuale, legato allo stato della natura e decidere, sulla base del risultato, se l'ipotesi può essere accettata o respinta. Accettare un'ipotesi non significa la stessa cosa per tutti e in tutte le situazioni. Sicuramente **non** significa che è stata provata la sua veridicità in modo rigoroso, perchè i dati di un campione non contengono tutta l'informazione sulla popolazione. "Accettare" vuol dire "credere" in senso debole. Vuol dire cioè comportarsi come se l'ipotesi fosse vera.

Il TEST di un'ipotesi È UNA REGOLA che assegna una delle etichette: "accettare l'ipotesi" o "respingere l'ipotesi" al risultato di un esperimento.

Un test è normalmente descritto in termini di qualche statistica $T = t(\mathbf{X})$. L'intervallo dei valori di T per i quali l'ipotesi è **respinta** viene detto **regione critica** del test. L'ipotesi da testare viene detta **ipotesi nulla**, mentre l'insieme degli altri stati della natura considerati possibili è detto **ipotesi alternativa**.

L'ipotesi nulla è indicata con H_0 ; l'altra con H_A o H_1 . L'ipotesi nulla *spesso* è l'affermazione di uguaglianza di due stati di natura, che deve essere provata. La prova può però portare a conclusioni sbagliate. Supponendo di utilizzare un test adatto (un buon test), questo può tuttavia portare a due tipi di errori:

I tipo: respingere H_0 quando H_0 è vera.

II tipo: accettare H_0 quando H_0 è falsa.

Le *dimensioni* di questi errori sono definite come:

$\alpha = P_{H_0}$ (respingere H_0)= dimensione dell'errore di I tipo

$\beta = P_{H_1}$ (accettare H_0)= dimensione dell'errore di II tipo

Normalmente l'errore di I tipo viene considerato il più grave; si fissa allora un valore accettabile per α e si minimizza β (i due errori sono strettamente legati). Ovviamente la scelta di α è soggettiva. La quantità $(1 - \alpha)$ è detta livello di confidenza e significa che *si è confidenti* che l'ipotesi è *non respinta* con $\alpha\%$ di probabilità di sbagliare (o con $(1 - \alpha)\%$ di probabilità di fare la scelta giusta).

Bisogna sottolineare il fatto che un test di ipotesi non contiene nessuna capacità di accettare un'ipotesi, ma solo di respingerla. Se un'ipotesi non si può respingere, allora la statistica non può dire nulla. Il fatto che abitualmente si accetti un'ipotesi è solo un'abitudine sbagliata.

TEST DI NORMALITÀ (confronto tra due medie)

Si abbiano due valori medi, X_1 e X_2 , di varianze σ_1 e σ_2 , rispettivamente, che seguano la legge normale. Ci si propone di verificare se le due medie sono uguali. Questo test é stato utilizzato nelle prove di laboratorio per verificare se due valori della stessa grandezza, ottenuti con metodi diversi, sono compatibili tra loro. Sarà allora controllata l'ipotesi nulla $H_0: X_1 - X_2 = 0$.

La variabile di test, che deriva dal fatto che $(X_1 - X_2)$ ha distribuzione normale con varianza $(\sigma_1^2 + \sigma_2^2)$, sarà:

$$z^* = \frac{(X_1 - X_2)}{\sqrt{\sigma_1^2 + \sigma_2^2}}.$$

Fissato un livello di significatività α , se $z_\alpha < z^*$ l'ipotesi H_0 sarà respinta con livello confidenza $(1 - \alpha)$.

TEST DI STUDENT (confronto tra medie)

Consideriamo la variabile aleatoria X e un suo campione di grandezza N , con media \bar{x} e varianza della media $\sigma^2(\bar{x}) = \sigma^2(x)/N$. Dal teorema limite centrale sappiamo che la statistica

$$y = \frac{\bar{x} - \hat{x}}{\sigma(\bar{x})}$$

dove \hat{x} è la media della popolazione (valore di aspettazione della media di un campione grande), segue una distribuzione normale. Purtroppo però $\sigma(\bar{x})$ non è mai noto. Al suo posto si conosce la stima di $\sigma(x)$

$$s_x^2 = \frac{1}{N-1} \sum_{j=1}^N (x_j - \bar{x})^2.$$

Da $\sigma^2(\bar{x}) = \sigma^2(x)/N$ deriviamo:

$$s_{\bar{x}}^2 = \frac{1}{N(N-1)} \sum_{j=1}^N (x_j - \bar{x})^2.$$

Ci chiediamo: se nella $y = \dots$ sostituiamo $\sigma(\bar{x})$ con $s_{\bar{x}}$, quanto y si discosta da una gaussiana? Consideriamo la statistica, trasladando tutto in modo che sia $\hat{x} = 0$:

$$t = \frac{\bar{x}}{s_{\bar{x}}} = \frac{\bar{x}\sqrt{N}}{s_x}.$$

La v.a. t segue la distribuzione di Student.

Per calcolare i limiti t_α corrispondenti ad un fissato livello di confidenza α , si utilizzano gli integrali:

$$\int_{-\infty}^{t_\alpha} f(t)dt = 1 - \alpha \quad \text{oppure} \quad \int_0^{t_\alpha} f(t)dt = \frac{1}{2} - \alpha.$$

Se un'ipotesi predice un valore di aspettazione a e da un campione di dimensione N si derivano \bar{x} e s_x^2 e se la disuguaglianza

$$|t| = \frac{|\bar{x} - a|\sqrt{N}}{s_x} > t_\alpha$$

è soddisfatta, l'ipotesi è respinta con livello di significatività α [o con livello di confidenza $(1 - \alpha)$].

Il test di Student (e la distribuzione relativa, t) è stato formulato da W. Gosset all'inizio del 1900. L'impiego di Gosset (come chimico nella fabbrica di birra Guinness a Dublino, dal 1899) comprendeva il segreto industriale, il che gli impedì di pubblicare la sua scoperta: ci si rese conto, però, dell'importanza della distribuzione t (usata alla Guinness per il controllo qualità da piccoli campioni) e a Gosset fu permesso di pubblicare sotto lo pseudonimo di "Student".

Gosset fece le scuole secondarie a Winchester e poi studiò chimica e matematica al New College di Oxford. In questo periodo uno dei suoi professori fu Airy.

Il test di Student può essere generalizzato per confrontare le medie di due campioni: queste medie hanno varianze $\sigma_{\bar{x}}^2 = \sigma^2(x)/N_1$; $\sigma_{\bar{y}}^2 = \sigma^2(y)/N_2$ e stime

$$s_{\bar{x}}^2 = \frac{1}{N_1(N_1-1)} \sum_{j=1}^{N_1} (x_j - \bar{x})^2$$

e

$$s_{\bar{y}}^2 = \frac{1}{N_2(N_2-1)} \sum_{j=1}^{N_2} (y_j - \bar{y})^2$$

La differenza delle medie $\Delta = \bar{x} - \bar{y}$ è distribuita normalmente con varianza $s_{\Delta}^2 = s_x^2 + s_y^2$. Se è vera l'ipotesi per cui le medie della popolazione sono uguali, il rapporto $\Delta/\sigma(\Delta)$ segue una gaussiana standardizzata. Potremmo quindi scrivere la probabilità di veridicità dell'ipotesi, se $\sigma(\Delta)$ fosse nota. Invece conosciamo s_{Δ} (la sua stima), e il rapporto Δ/s_{Δ} non segue una distribuzione gaussiana, ma una distribuzione di Student. L'ipotesi $\hat{x} = \hat{y}$ implica che \bar{x} e \bar{y} derivano dalla stessa popolazione e che $\sigma^2(x) = \sigma^2(y)$. Quindi la migliore stima (s^2) di s_x^2 e s_y^2 è la media, pesata con i gradi di libertà. Allora

$$s^2 = \frac{(N_1 - 1)s_x^2 + (N_2 - 1)s_y^2}{(N_1 - 1) + (N_2 - 1)}$$

e

$$s_{\Delta}^2 = \frac{N_1 + N_2}{N_1 N_2} s^2.$$

La Δ/s_{Δ} segue una distribuzione di Student con $(N_1 + N_2 - 2)$ gradi di libertà, e se

$$|t| = \frac{|\Delta|}{s_{\Delta}} = \frac{|\bar{x} - \bar{y}|}{s_{\Delta}} > t_{\alpha}$$

l'ipotesi di medie uguali è respinta con livello di significatività α .

TEST PER LE MEDIE (descrizione alternativa che comprende il test di Student)

1) Varianza della popolazione nota

Consideriamo un campione casuale X_1, X_2, \dots, X_n di dimensione n , estratto da una popolazione normale di cui è nota la varianza σ^2 .

Per verificare l'ipotesi che il campione provenga da una popolazione con media (valore di aspettazione) $\hat{X} = \mu$ si usa la statistica;

$$Z = \frac{(\bar{X} - \mu)\sqrt{n}}{\sigma}$$

con \bar{X} media del campione e σ^2 varianza della popolazione, che segue la distribuzione normale.

La regione di rifiuto, per il **test a due code**, è data dalla disuguaglianza:

$$\frac{(\bar{X} - \mu)\sqrt{n}}{\sigma} > z_{1-\frac{\alpha}{2}} \text{ oppure } \frac{(\bar{X} - \mu)\sqrt{n}}{\sigma} < -z_{1-\frac{\alpha}{2}}$$

o, risolvendo rispetto a \bar{X} :

$$\bar{X} > z_{\frac{1-\alpha}{2}} \frac{\sigma}{\sqrt{n}} + \mu \text{ oppure } \bar{X} < -z_{\frac{1-\alpha}{2}} \frac{\sigma}{\sqrt{n}} + \mu$$

• Se il test è a **una** coda, cioè se si vuole testare l'ipotesi $H_o : \hat{\mathbf{X}} = \mu$ contro l'ipotesi alternativa $H_o : \hat{\mathbf{X}} > \mu$, ad esempio, la regione di rifiuto sarà definita da:

$$\frac{(\bar{X}-\mu)\sqrt{n}}{\sigma} > z_{1-\alpha} \text{ oppure } \frac{(\bar{X}-\mu)\sqrt{n}}{\sigma} < -z_{1-\alpha} \text{ se è } H_o : \hat{\mathbf{X}} < \mu$$

2) Varianza della popolazione non nota

Consideriamo un campione casuale X_1, X_2, \dots, X_n di dimensione n , estratto da una popolazione normale.

Per verificare l'ipotesi che il campione provenga da una popolazione con media (valore di aspettazione) $\hat{\mathbf{X}} = \mu$, nel caso in cui la varianza della popolazione non sia nota, si usa la statistica;

$$T = \frac{(\bar{X} - \mu)\sqrt{n}}{s}$$

con \bar{X} media del campione e s^2 varianza del campione, che segue la distribuzione t di Student.

La regione di rifiuto, per il **test a due code**, è data dalla disuguaglianza:

$$\frac{(\bar{X}-\mu)\sqrt{n}}{s} > t_{\nu, 1-\frac{\alpha}{2}} \text{ oppure } \frac{(\bar{X}-\mu)\sqrt{n}}{s} < -t_{\nu, 1-\frac{\alpha}{2}}$$

o, risolvendo rispetto a \bar{X} :

$$\bar{X} > t_{\nu, 1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} + \mu \text{ oppure } \bar{X} < -t_{\nu, 1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} + \mu$$

TEST PER LA DIFFERENZA TRA DUE MEDIE (descrizione alternativa che comprende il test di Student)

1) Varianze σ_1^2 e σ_2^2 delle popolazioni note

Se si vuole testare l'ipotesi di uguaglianza dei valori di aspettazione delle popolazioni da cui sarebbero estratti i due campioni, di dimensioni n_1 e n_2 e di medie \bar{X}_1 e \bar{X}_2 , cioè $H_o : \mu_1 - \mu_2 = 0$ si usa la statistica

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sigma_{\Delta}}; \quad \sigma_{\Delta}^2 = \sigma_{\bar{X}_1 - \bar{X}_2}^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

La regione di rifiuto di H_o sarà data da (**2 code**)

$$\frac{\bar{X}_1 - \bar{X}_2}{\sigma_{\Delta}} > z_{1-\frac{\alpha}{2}} \text{ oppure } \frac{\bar{X}_1 - \bar{X}_2}{\sigma_{\Delta}} < -z_{1-\frac{\alpha}{2}}$$

Per il test a **1 coda**, in cui, $H_1 : \mu_1 - \mu_2 > 0$, la regione di rifiuto sarà data da:

$$\frac{\bar{X}_1 - \bar{X}_2}{\sigma_{\Delta}} > z_{1-\alpha}$$

e, quando è $H_1 : \mu_1 - \mu_2 < 0$, da

$$\frac{\bar{X}_1 - \bar{X}_2}{\sigma_{\Delta}} < -z_{1-\alpha}$$

2) Varianze σ_1^2 e σ_2^2 non note, ma ritenute uguali

Avendo a disposizione due campioni, di dimensione n_1 e n_2 , che forniscono medie \bar{X}_1 e \bar{X}_2 e varianze campionarie s_1^2 e s_2^2 , vogliamo testare l'ipotesi nulla $H_0 : \mu_1 - \mu_2 = 0$ contro l'ipotesi alternativa $H_1 : \mu_1 - \mu_2 \neq 0$. Siano μ_1 e μ_2 i valori di aspettazione delle popolazioni da cui sarebbero stati estratti i campioni.

Si usa in questo caso lo stimatore:

$$T = \frac{\bar{X}_1 - \bar{X}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}; \quad S^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

che segue la distribuzione di Student con $(n_1 + n_2 - 2)$ gradi di libertà.

L'ipotesi nulla viene rifiutata se:

$$\frac{\bar{X}_1 - \bar{X}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} > t_{n_1+n_2-2, \frac{1-\alpha}{2}}$$

oppure se

$$\frac{\bar{X}_1 - \bar{X}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} < -t_{n_1+n_2-2, \frac{1-\alpha}{2}}$$

ESERCIZIO:

Verificare se il campione 19.4, 22.3, 21.2, 22, 23.7, 20 è compatibile con una gaussiana di media 22.8, con livello di significatività del 10 %.

Soluzione

La media e la varianza del campione ($N = 6$; $\nu = 5$) sono $\bar{x} = 21.43$ e $s_x^2 = 2.50$. quindi

$$|t| = \frac{|21.43 - 22.8|}{s/\sqrt{6}} = \frac{|-1.37|}{1.58} 2.45 = 2.12$$

Questo è un test a due code, quindi la significatività del 10% va divisa per le due code ($1 - \alpha/2 = 0.95$).

Dalle tabelle si legge, per $\nu = 5$, $t_\alpha = 2.02$. Allora $|t| > t_\alpha$ e l'ipotesi *deve* essere respinta. Il campione non è quindi compatibile, al livello di confidenza del 90%, con la gaussiana data.

F-TEST SULL'UGUAGLIANZA DELLE VARIANZE (FISHER)

Consideriamo due campioni di grandezza N_1 e N_2 . Ci chiediamo se essi derivino dalla stessa popolazione o no. Sappiamo che i due campioni hanno la stessa media e varianze empiriche diverse.

Dobbiamo quindi testare l'ipotesi zero $H_0 = \sigma_1^2 - \sigma_2^2 = 0$.

Per ognuno dei campioni calcoliamo le varianze empiriche s_1^2 e s_2^2 e poi la statistica

$$F = \frac{s_1^2}{s_2^2}$$

Sappiamo che le quantità

$$R_1^2 = \frac{\nu_1 s_1^2}{\sigma_1^2} \quad e \quad R_2^2 = \frac{\nu_2 s_2^2}{\sigma_2^2} \quad (v. pag. 79, def. di U),$$

con ν gradi di libertà e σ varianza della popolazione, seguono una legge χ^2 con ν_1 e ν_2 gradi di libertà. Se l'ipotesi è vera si ha

$$F = \frac{\nu_2 R_1^2}{\nu_1 R_2^2}.$$

Utilizzando la χ^2 per R_1^2 e R_2^2 si vede che la densità di probabilità per F è la distribuzione di Fisher.

IN PRATICA: dati i campioni con ν_1 e ν_2 , si calcola il rapporto s_1^2/s_2^2 . Questo numero, per ν_1 e ν_2 , si confronta con le tavole, dato α . Se $s_1^2/s_2^2 > F_{tabulato}$ diremo che l'ipotesi $\sigma_1^2 = \sigma_2^2$ è rigettata al livello di confidenza $(1 - \alpha)$. Se $s_1^2/s_2^2 < F_{tabulato}$ diremo che l'ipotesi $\sigma_1^2 = \sigma_2^2$ non può essere respinta, al livello di confidenza $(1 - \alpha)$.

Sir Ronald A. Fisher prese un diploma in Astronomia a Cambridge nel 1912. Là studiò la teoria degli errori usando il manuale di Airy "Teoria degli errori". Il suo interesse per la teoria degli errori in Astronomia lo portò ad approfondire i problemi statistici. Smise di insegnare matematica per andare nel 1919 a lavorare in una stazione agricola sperimentale come biologo: diede molti contributi sia in statistica che in genetica. Successivamente introdusse il concetto di analisi della varianza. Il contributo di Fisher riguarda lo sviluppo di metodi adatti a piccoli campioni (come Gosset), la scoperta delle distribuzioni di molti campioni statistici e l'analisi della varianza. Introdusse il termine **massima verosimiglianza** e studiò il test delle ipotesi.

Fisher viene considerato uno dei padri della statistica moderna.

ESEMPIO: Dati i due campioni, con 8 e 6 gradi di libertà, rispettivamente,

- 1) 21,19,14,27,25,23,22,18,21
- 2) 16,24,22,21,25,21,18

si vuole verificare l'ipotesi che il campione 2) abbia varianza minore del campione 1), al livello di significatività $\alpha = 5\%$

Soluzione: è $\nu_1 = 8$, $\nu_2 = 6$; $s_1^2 = 14.91$, $s_2^2 = 10$ e $F_{sperim.} = s_1^2/s_2^2 = 1.491$; $F_{0.95}(8, 6) = 4.15$. Allora, essendo $F_{sperim.} < F_{tabulato}$ diremo che s_2^2 non è significativamente minore di s_1^2 .

TEST DI χ^2 - ADATTAMENTO DI CURVE

Questo test si usa per confrontare la funzione di distribuzione di un campione con la distribuzione della popolazione (distribuzione parente), assunta in seguito ad una ipotesi e viene normalmente usato per verificare la bontà di un fit.

Consideriamo una variabile aleatoria X con funzione di distribuzione $F(x)$ e densità $f(x)$. L'intervallo di definizione della X può essere diviso in r intervalli $\xi_1, \xi_2 \dots \xi_r$.

La probabilità di osservare X nel ξ -esimo intervallo è:

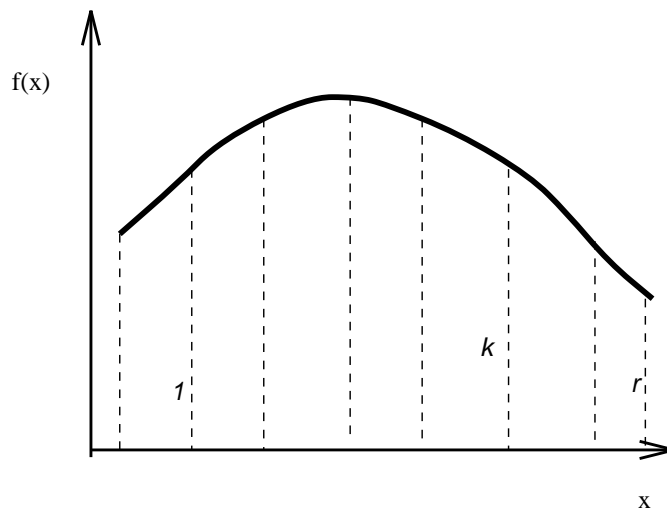
$$p_k = P(x \in \xi_k) = \int_{\xi_k} f(x) dx \quad \sum_{k=1}^r p_k = 1.$$

Estraiamo un campione di dimensione n e chiamiamo n_k il numero di elementi del campione che cadono nell'intervallo ξ_k . Sarà evidentemente $\sum_{k=1}^r n_k = n$. Dalla densità di probabilità della *popolazione* ci aspettiamo che $n_k = np_k$.

Come misura della deviazione dei dati del campione dalla previsione teorica usiamo la statistica

$$X^2 = \sum_{k=1}^r \frac{(n_k - np_k)^2}{np_k}$$

(percentuale della deviazione quadratica). X^2 segue, per $n \rightarrow \infty$, una distribuzione χ^2 con $r - 1$ gradi di libertà. Se devono essere calcolati p parametri, allora $\nu = r - p - 1$.



IN PRATICA: dai dati sperimentali e dalla legge teorica si costruisce

$$X^2 = \sum_{i=1}^n \frac{(x_{i,sperim} - x_{i,teor})^2}{x_{i,teor}};$$

si tratta poi questa variabile come una χ^2 con $n-1$ o $n-p-1$ gradi di libertà. Se, fissato $1-\alpha$, $X^2 > \chi^2$ l'ipotesi (l'uguaglianza tra le distribuzioni osservata e teorica) è respinta con livello di significatività α .

Esercizio: Si supponga che il numero N di galassie per unità di superficie (sulla sfera celeste) sia una funzione del flusso luminoso S , secondo la legge:

$$N = S^a.$$

Nell'ipotesi che $a=-1.5$, verificare se i dati della successiva tabella seguono la legge data, al livello di significatività $\alpha = 0.01$.

S	1	2	3	4	5	6	7
N	60	25	10	9	4	5	3

Soluzione

Dalla legge, i valori teorici (attesi), in frequenza e in numero assoluto, normalizzati al primo dei valori osservati, sono:

S	1	2	3	4	5	6	7
f	1	0.35	0.19	0.125	0.09	0.07	0.05
N*	60	21	11.4	7.5	5.4	4.2	3

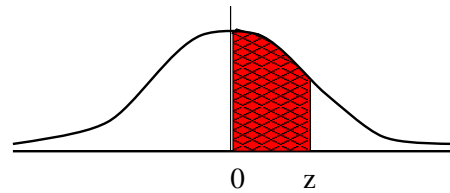
Si può allora calcolare X^2 da N e N^* :

$$X^2 = \sum_{i=1}^7 \frac{(N - N^*)^2}{N} = \frac{(25 - 21)^2}{21} + \dots + \frac{(5 - 4.2)^2}{4.2} = 1.749.$$

Al livello di significatività $\alpha = 0.01$ ($\alpha/2 = 0.005$) risulta, per $\nu = 6$: $\chi_{0.005}^2 < X^2 < \chi_{0.995}^2$, da cui l'ipotesi

H_0 : "i dati seguono la legge, con parametro $a=-1.5$ " non può essere respinta, al livello di confidenza del 99

**Aree sotto la curva
normale standard
tra 0 e z**



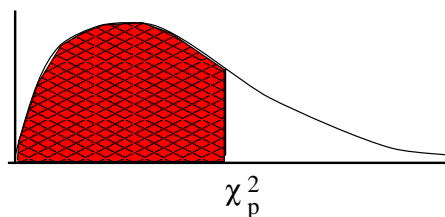
z	0	1	2	3	4	5	6	7	8	9
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0754
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2258	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2518	.2549
0.7	.2580	.2612	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2996	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.5	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998
3.6	.4998	.4998	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.7	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.8	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.9	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000

**Quantili x_p della
distribuzione normale**

$$P = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x_p} \exp(-x^2/2) dx$$

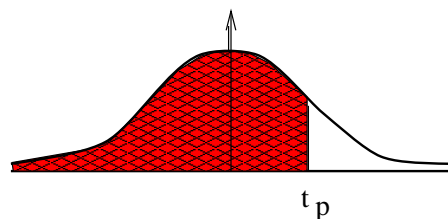
P	0	1	2	3	4	5	6	7	8	9
0.0	$-\infty$	-2.33	-2.05	-1.88	-1.75	-1.64	-1.55	-1.48	-1.41	-1.34
0.1	-1.28	-1.23	-1.17	-1.13	-1.08	-1.04	-0.99	-0.95	-0.92	-0.88
0.2	-0.84	-0.81	-0.77	-0.74	-0.71	-0.67	-0.64	-0.61	-0.58	-0.55
0.3	-0.52	-0.50	-0.47	-0.44	-0.41	-0.39	-0.36	-0.33	-0.31	-0.28
0.4	-0.25	-0.23	-0.20	-0.18	-0.15	-0.13	-0.10	-0.08	-0.05	-0.03
0.5	0.00	0.03	0.05	0.08	0.10	0.13	0.15	0.18	0.20	0.23
0.6	0.25	0.28	0.31	0.33	0.36	0.39	0.41	0.44	0.47	0.50
0.7	0.52	0.55	0.58	0.61	0.64	0.67	0.71	0.74	0.77	0.81
0.8	0.84	0.88	0.92	0.95	0.99	1.04	1.08	1.13	1.17	1.23
0.9	1.28	1.34	1.41	1.48	1.55	1.64	1.75	1.88	2.05	2.33

**Valori Percentuali $\chi^2(p)$
per la distribuzione χ^2
con ν gradi di libertà
(area tratteggiata = p)**



ν	.995	.99	.975	.95	.90	.75	.50	.25	.10	.05	.025	.01	.005
1	7.88	6.63	5.02	3.84	2.71	1.32	.455	.102	.0158	.0039	.0010	.0002	.0000
2	10.6	9.21	7.38	5.99	4.61	2.77	1.39	.575	.211	.103	.0506	.0201	.0100
3	12.8	11.3	9.35	7.81	6.25	4.11	2.37	1.21	.584	.352	.216	.115	.0717
4	14.9	13.3	11.1	9.49	7.78	5.39	3.36	1.92	1.06	.711	.484	.297	.207
5	16.7	15.1	12.8	11.1	9.24	6.63	4.35	2.67	1.61	1.15	.831	.554	.412
6	18.5	16.8	14.4	12.6	10.6	7.84	5.35	3.45	2.20	1.64	1.24	.872	.676
7	20.3	18.5	16.0	14.1	12.0	9.04	6.35	4.25	2.83	2.17	1.69	1.24	.989
8	22.0	20.1	17.5	15.5	13.4	10.2	7.34	5.07	3.49	2.73	2.18	1.65	1.34
9	23.6	21.7	19.0	16.9	14.7	11.4	8.34	5.90	4.17	3.33	2.70	2.09	1.73
10	25.2	23.2	20.5	18.3	16.0	12.5	9.34	6.74	4.87	3.94	3.25	2.56	2.16
11	26.8	24.7	21.9	19.7	17.3	13.7	10.3	7.58	5.58	4.57	3.82	3.05	2.60
12	28.3	26.2	23.3	21.0	18.5	14.8	11.3	8.44	6.30	5.23	4.40	3.57	3.07
13	29.8	27.7	24.7	22.4	19.8	16.0	12.3	9.30	7.04	5.89	5.01	4.11	3.57
14	31.3	29.1	26.1	23.7	21.1	17.1	13.3	10.2	7.79	6.57	5.63	4.66	4.07
15	32.8	30.6	27.5	25.0	22.3	18.2	14.3	11.0	8.55	7.26	6.26	5.23	4.60
16	34.3	32.0	28.8	26.3	23.5	19.4	15.3	11.9	9.31	7.96	6.91	5.81	5.14
17	35.7	33.4	30.2	27.6	24.8	20.5	16.3	12.8	10.1	8.67	7.56	6.41	5.70
18	37.2	34.8	31.5	28.9	26.0	21.6	17.3	13.7	10.9	9.39	8.23	7.01	6.26
19	38.6	36.2	32.9	30.1	27.2	22.7	18.3	14.6	11.7	10.1	8.91	7.63	6.84
20	40.0	37.6	34.2	31.4	28.4	23.8	19.3	15.5	12.4	10.9	9.59	8.26	7.43
21	41.4	38.9	35.5	32.7	29.6	24.9	20.3	16.3	13.2	11.6	10.3	8.90	8.03
22	42.8	40.3	36.8	33.9	30.8	26.0	21.3	17.2	14.0	12.3	11.0	9.54	8.64
23	44.2	41.6	38.1	35.2	32.0	27.1	22.3	18.1	14.8	13.1	11.7	10.2	9.26
24	45.6	43.0	39.4	36.4	33.2	28.2	23.3	19.0	15.7	13.8	12.4	10.9	9.89
25	46.9	44.3	40.6	37.7	34.4	29.3	24.3	19.9	16.5	14.6	13.1	11.5	10.5
26	48.3	45.6	41.9	38.9	35.6	30.4	25.3	20.8	17.3	15.4	13.8	12.2	11.2
27	49.6	47.0	43.2	40.1	36.7	31.5	26.3	21.7	18.1	16.2	14.6	12.9	11.8
28	51.0	48.3	44.5	41.3	37.9	32.6	27.3	22.7	18.9	16.9	15.3	13.6	12.5
29	52.3	49.6	45.7	42.6	39.1	33.7	28.3	23.6	19.8	17.7	16.0	14.3	13.1
30	53.7	50.9	47.0	43.8	40.3	34.8	29.3	24.5	20.6	18.5	16.8	15.0	13.8
40	66.8	63.7	59.3	55.8	51.8	45.6	39.3	33.7	29.1	26.5	24.4	22.2	20.7
50	79.5	76.2	71.4	67.5	63.2	56.3	49.3	42.9	37.7	34.8	32.4	29.7	28.0
60	92.0	88.4	83.3	79.1	74.4	67.0	59.3	52.3	46.5	43.2	40.5	37.5	35.5
70	104.2	100.4	95.0	90.5	85.5	77.6	69.3	61.7	55.3	51.7	48.8	45.4	43.3
80	116.3	112.3	106.6	101.9	96.6	88.1	79.3	71.1	64.3	60.4	57.2	53.5	51.2
90	128.3	124.1	118.1	113.1	107.6	98.6	89.3	80.6	73.3	69.1	65.6	61.8	59.2
100	140.2	135.8	129.6	124.3	118.5	109.1	99.3	90.1	82.4	77.9	74.2	70.1	67.3

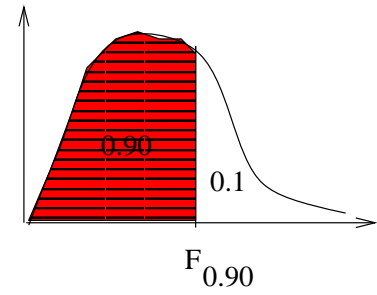
Valori percentuali (t_p) per
la distribuzione di Student
con ν gradi di libertà
(area tratteggiata = p)



ν	.995	.99	.975	.95	.90	.80	.75	.70	.60	.55
1	63.66	31.82	12.71	6.31	3.08	1.376	1.000	.727	.325	.158
2	9.92	6.96	4.30	2.92	1.89	1.061	.816	.617	.289	.142
3	5.84	4.54	3.18	2.35	1.64	.978	.765	.584	.277	.137
4	4.60	3.75	2.78	2.13	1.53	.941	.741	.569	.271	.134
5	4.03	3.36	2.57	2.02	1.48	.920	.727	.559	.267	.132
6	3.71	3.14	2.45	1.94	1.44	.906	.718	.553	.265	.131
7	3.50	3.00	2.36	1.90	1.42	.896	.711	.549	.263	.130
8	3.36	2.90	2.31	1.86	1.40	.889	.706	.546	.262	.130
9	3.25	2.82	2.26	1.83	1.38	.883	.703	.543	.261	.129
10	3.17	2.76	2.23	1.81	1.37	.879	.700	.542	.260	.129
11	3.11	2.72	2.20	1.80	1.36	.876	.697	.540	.260	.129
12	3.06	2.68	2.18	1.78	1.36	.873	.695	.539	.259	.128
13	3.01	2.65	2.16	1.77	1.35	.870	.694	.538	.259	.128
14	2.98	2.62	2.14	1.76	1.34	.868	.692	.537	.258	.128
15	2.95	2.60	2.13	1.75	1.34	.866	.691	.536	.258	.128
16	2.92	2.58	2.12	1.75	1.34	.865	.690	.535	.258	.128
17	2.90	2.57	2.11	1.74	1.33	.863	.689	.534	.257	.128
18	2.88	2.55	2.10	1.73	1.33	.862	.688	.534	.257	.127
19	2.86	2.54	2.09	1.73	1.33	.861	.688	.533	.257	.127
20	2.84	2.53	2.09	1.72	1.32	.860	.687	.533	.257	.127
21	2.83	2.52	2.08	1.72	1.32	.859	.686	.532	.257	.127
22	2.82	2.51	2.07	1.72	1.32	.858	.686	.532	.256	.127
23	2.81	2.50	2.07	1.71	1.32	.858	.685	.532	.256	.127
24	2.80	2.49	2.06	1.71	1.32	.857	.685	.531	.256	.127
25	2.79	2.48	2.06	1.71	1.32	.856	.684	.531	.256	.127
26	2.78	2.48	2.06	1.71	1.32	.856	.684	.531	.256	.127
27	2.77	2.47	2.05	1.70	1.31	.855	.684	.531	.256	.127
28	2.76	2.47	2.05	1.70	1.31	.855	.683	.530	.256	.127
29	2.76	2.46	2.04	1.70	1.31	.854	.683	.530	.256	.127
30	2.75	2.46	2.04	1.70	1.31	.854	.683	.530	.256	.127
40	2.70	2.42	2.02	1.68	1.30	.851	.681	.529	.255	.126
60	2.66	2.39	2.00	1.67	1.30	.848	.679	.527	.254	.126
120	2.62	2.36	1.98	1.66	1.29	.845	.677	.526	.254	.126
∞	2.58	2.33	1.96	1.645	1.28	.842	.674	.524	.253	.126

Distribuzione di FISHER.
Valori dei frattili $F_p(\nu_1, \nu_2)$
 $[\nu_1, \nu_2$ gradi di libertà dei
campioni], dati da:

$$P = \int_0^{F_p} f(F) dF = 0.90$$



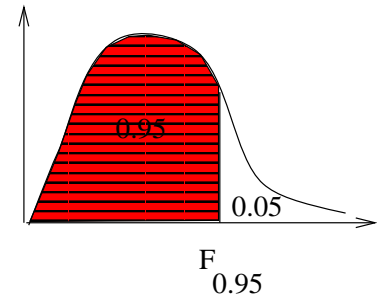
ν_2	1	2	3	4	5	6	8	10	12	14	16	20	30	40	50	∞
1	39.86	49.50	53.59	55.83	57.24	58.20	59.44	60.19	60.70	61.07	61.35	61.74	62.26	62.53	62.70	63.33
2	8.53	9.00	9.16	9.24	9.29	9.33	9.37	9.39	9.41	9.42	9.43	9.44	9.46	9.47	9.47	9.49
3	5.54	5.46	5.39	5.34	5.31	5.29	5.25	5.23	5.22	5.20	5.20	5.18	5.17	5.16	5.15	5.13
4	4.54	4.32	4.19	4.11	4.05	4.01	3.96	3.92	3.90	3.88	3.86	3.84	3.82	3.80	3.80	3.76
5	4.06	3.78	3.62	3.52	3.45	3.40	3.34	3.30	3.27	3.25	3.23	3.21	3.17	3.16	3.15	3.10
6	3.78	3.46	3.29	3.18	3.11	3.05	2.98	2.94	2.90	2.88	2.86	2.84	2.80	2.78	2.77	2.72
7	3.59	3.26	3.07	2.96	2.88	2.83	2.75	2.70	2.67	2.64	2.62	2.59	2.56	2.54	2.52	2.47
8	3.46	3.11	2.92	2.81	2.73	2.67	2.59	2.54	2.50	2.48	2.45	2.42	2.38	2.36	2.35	2.29
9	3.36	3.01	2.81	2.69	2.61	2.55	2.47	2.42	2.38	2.35	2.33	2.30	2.25	2.23	2.22	2.16
10	3.29	2.92	2.73	2.61	2.52	2.46	2.38	2.32	2.28	2.26	2.23	2.20	2.16	2.13	2.12	2.06
11	3.23	2.86	2.66	2.54	2.45	2.39	2.30	2.25	2.21	2.18	2.16	2.12	2.08	2.05	2.04	1.97
12	3.18	2.81	2.61	2.48	2.39	2.33	2.24	2.19	2.15	2.12	2.09	2.06	2.01	1.99	1.97	1.90
13	3.14	2.76	2.56	2.43	2.35	2.28	2.20	2.14	2.10	2.07	2.04	2.01	1.96	1.93	1.92	1.85
14	3.10	2.73	2.52	2.39	2.31	2.24	2.15	2.10	2.05	2.02	2.00	1.96	1.91	1.89	1.87	1.80
15	3.07	2.70	2.49	2.36	2.27	2.21	2.12	2.06	2.02	1.99	1.96	1.92	1.87	1.85	1.83	1.76
16	3.05	2.67	2.46	2.33	2.24	2.18	2.09	2.03	1.99	1.95	1.93	1.89	1.84	1.81	1.79	1.72
17	3.03	2.64	2.44	2.31	2.22	2.15	2.06	2.00	1.96	1.93	1.90	1.86	1.81	1.78	1.76	1.69
18	3.01	2.62	2.42	2.29	2.20	2.13	2.04	1.98	1.93	1.90	1.87	1.84	1.78	1.75	1.74	1.66
19	2.99	2.61	2.40	2.27	2.18	2.11	2.02	1.96	1.91	1.88	1.85	1.81	1.76	1.73	1.71	1.63
20	2.91	2.59	2.38	2.25	2.16	2.09	2.00	1.94	1.89	1.86	1.83	1.79	1.74	1.71	1.69	1.61
25	2.92	2.53	2.32	2.18	2.09	2.02	1.93	1.87	1.82	1.79	1.76	1.72	1.66	1.63	1.61	1.52
30	2.88	2.49	2.28	2.14	2.05	1.98	1.88	1.82	1.77	1.74	1.71	1.67	1.61	1.51	1.55	1.46
35	2.85	2.46	2.25	2.11	2.02	1.95	1.85	1.79	1.74	1.70	1.67	1.63	1.57	1.53	1.51	1.41
40	2.84	2.44	2.23	2.09	2.00	1.93	1.83	1.76	1.71	1.68	1.65	1.61	1.54	1.51	1.48	1.38
45	2.82	2.42	2.21	2.07	1.98	1.91	1.81	1.74	1.70	1.66	1.63	1.58	1.52	1.48	1.46	1.35
50	2.81	2.41	2.20	2.06	1.97	1.90	1.80	1.73	1.68	1.64	1.61	1.57	1.50	1.46	1.44	1.33
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.67	1.60	1.55	1.50	1.47	1.42	1.34	1.30	1.26	1.00

I valori $F_{0.05}, F_{0.025}, F_{0.005}$ si possono calcolare tramite la formula:

$$F_{\alpha}(\nu_1, \nu_2) = \frac{1}{F_{1-\alpha}(\nu_2, \nu_1)}$$

Distribuzione di FISHER.
Valori dei frattili $F_p(\nu_1, \nu_2)$
 [ν_1, ν_2 gradi di libertà dei
 campioni], dati da:

$$P = \int_0^{F_p} f(F) dF = 0.95$$



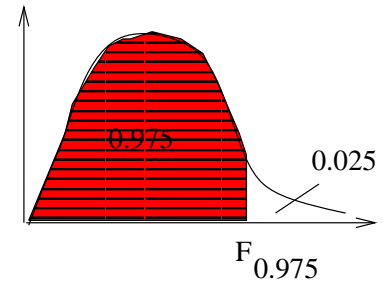
ν_2	1	2	3	4	5	6	8	10	12	14	16	20	30	40	50	∞
1	161.4	199.5	215.6	224.5	230.2	234.0	238.9	241.9	243.9	245.4	246.5	248.0	250.1	251.0	252.0	254.4
2	18.51	19.00	19.16	19.25	19.30	19.33	19.37	19.40	19.41	19.42	19.43	19.44	19.46	19.47	19.47	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.85	8.79	8.75	8.71	8.69	8.66	8.62	8.59	8.58	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.04	5.96	5.91	5.81	5.84	5.80	5.75	5.72	5.70	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.82	4.74	4.68	4.64	4.60	4.56	4.50	4.46	4.44	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.15	4.06	4.00	3.96	3.92	3.87	3.81	3.77	3.75	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.73	3.64	3.57	3.53	3.49	3.44	3.38	3.34	3.32	3.23
8	5.32	4.46	4.07	3.84	3.89	3.58	3.44	3.35	3.28	3.24	3.20	3.15	3.08	3.04	3.02	2.93
9	5.12	4.26	2.86	3.62	3.48	3.37	3.23	3.14	3.07	3.03	2.99	2.94	2.86	2.83	2.80	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.07	2.98	2.91	2.86	2.83	2.77	2.70	2.66	2.64	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	2.95	2.85	2.79	2.74	2.10	2.65	2.57	2.53	2.51	2.40
12	4.75	3.89	2.49	3.26	3.11	3.00	2.85	2.75	2.69	2.64	2.60	2.54	2.47	2.43	2.40	2.30
13	4.67	3.81	3.41	3.18	2.03	2.92	2.77	2.67	2.60	2.55	2.51	2.46	2.38	2.34	2.31	2.21
14	4.80	3.74	3.34	3.11	2.96	2.85	2.70	2.60	2.53	2.48	2.44	2.39	2.31	2.27	2.24	2.13
15	4.54	3.68	2.29	3.06	2.90	2.79	2.64	2.54	2.48	2.42	2.38	2.33	2.25	2.20	2.18	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.59	2.49	2.42	2.37	2.33	2.28	2.19	2.15	2.12	2.01
17	4.45	3.59	2.20	2.96	2.81	2.70	2.55	2.45	2.38	2.33	2.29	2.23	2.15	2.10	2.08	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.51	2.41	2.34	2.29	2.25	2.19	2.11	2.06	2.04	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.48	2.38	2.31	2.26	2.21	2.16	2.07	2.03	2.00	1.58
20	4.35	3.49	3.10	2.87	2.71	2.60	2.45	2.35	2.28	2.22	2.18	2.12	2.04	1.99	1.97	1.84
25	4.24	3.39	2.99	2.76	2.60	2.49	2.34	2.24	2.16	2.11	2.07	2.01	1.92	1.87	1.84	1.71
30	4.17	3.32	2.92	2.69	2.53	2.42	2.27	2.16	2.09	2.04	1.99	1.93	1.84	1.79	1.76	1.62
35	4.12	3.27	2.87	2.64	2.49	2.37	2.22	2.11	2.04	1.99	1.94	1.88	1.79	1.74	1.70	1.56
40	4.08	3.23	2.84	2.61	2.45	2.34	2.18	2.08	2.00	1.95	1.90	1.84	1.74	1.69	1.66	1.51
45	4.06	3.20	2.81	2.58	2.42	2.31	2.15	2.05	1.97	1.92	1.87	1.81	1.71	1.66	1.63	1.47
50	4.03	3.18	2.79	2.56	2.40	2.29	2.13	2.03	1.95	1.89	1.85	1.78	1.69	1.63	1.60	1.44
∞	3.84	3.00	2.80	2.37	2.21	2.10	1.94	1.83	1.75	1.69	1.64	1.57	1.46	1.39	1.35	1.00

I valori $F_{0.05}, F_{0.025}, F_{0.005}$ si possono calcolare tramite la formula:

$$F_\alpha(\nu_1, \nu_2) = \frac{1}{F_{1-\alpha}(\nu_2, \nu_1)}$$

Distribuzione di FISHER.
Valori dei frattili $F_p(\nu_1, \nu_2)$
[ν_1, ν_2 gradi di libertà dei
campioni], dati da:

$$P = \int_0^{F_p} f(F) dF = 0.975$$



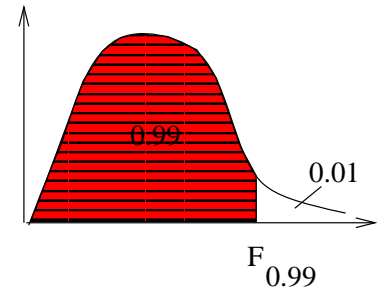
ν_2	ν_1															
	1	2	3	4	5	6	8	10	12	14	16	20	30	40	50	∞
1	647.8	799.4	864.0	399.4	921.6	936.8	956.3	968.3	976.3	982.1	986.5	992.7	1001	1006	1008	1018
2	38.50	39.00	39.17	39.25	39.30	39.33	39.37	39.39	39.41	39.42	39.43	39.44	39.46	39.47	39.48	39.50
3	17.44	16.04	15.44	15.10	14.88	14.73	14.54	14.42	14.34	14.28	14.23	14.17	14.08	14.04	14.01	13.90
4	12.22	10.65	9.98	9.60	9.36	9.20	8.98	8.84	8.75	8.68	8.63	8.56	8.46	8.41	8.38	8.26
5	10.01	8.43	7.76	7.39	7.15	6.98	6.76	6.62	6.52	6.46	6.40	6.33	6.23	6.18	6.14	6.02
6	8.81	7.26	6.60	6.23	5.99	5.82	5.60	5.46	5.37	5.30	5.24	5.17	5.07	5.01	4.98	4.85
7	8.07	6.54	5.89	5.52	5.29	5.12	4.90	4.76	4.67	4.60	4.54	4.47	4.36	4.31	4.28	4.14
8	7.57	6.06	5.42	5.05	4.82	4.65	4.43	4.30	4.20	4.13	4.08	4.00	3.89	3.84	3.81	3.67
9	7.21	5.71	5.08	4.72	4.48	4.32	4.10	3.96	3.87	3.80	3.74	3.67	3.56	3.51	3.47	3.33
10	6.94	5.46	4.83	4.47	4.24	4.07	3.85	3.72	3.62	3.55	3.50	3.42	3.31	3.26	3.22	3.08
11	6.72	5.26	4.63	4.28	4.04	3.88	3.66	3.53	3.43	3.36	3.30	3.23	3.12	3.06	3.03	2.88
12	6.55	5.10	4.47	4.12	3.89	3.73	3.51	3.37	3.28	3.21	3.15	3.07	2.96	2.91	2.87	2.72
13	6.41	4.97	4.35	4.00	3.77	3.60	3.39	3.25	3.15	3.08	3.03	2.95	2.84	2.78	2.74	2.60
14	6.30	4.86	4.24	3.89	3.66	3.50	3.29	3.15	3.05	2.98	2.92	2.84	2.73	2.67	2.64	2.49
15	6.20	4.77	4.15	3.80	3.58	3.41	3.20	3.06	2.96	2.89	2.84	2.76	2.64	2.59	2.55	2.40
16	6.12	4.69	4.08	3.73	3.50	3.34	3.12	2.99	2.89	2.82	2.76	2.68	2.57	2.52	2.47	2.32
17	6.04	4.62	4.01	3.66	3.44	3.26	3.06	2.92	2.82	2.75	2.70	2.62	2.50	2.44	2.41	2.25
18	5.98	4.56	3.95	3.61	3.38	3.22	3.01	2.87	2.77	2.70	2.64	2.56	2.44	2.38	2.35	2.19
19	5.92	4.51	3.90	3.56	3.33	3.17	2.96	2.82	2.72	2.65	2.59	2.51	2.39	2.33	2.30	2.13
20	5.87	4.46	3.86	3.51	3.29	3.13	2.91	2.77	2.68	2.60	2.55	2.46	2.35	2.29	2.25	2.09
25	5.69	4.29	3.69	3.35	3.13	2.97	2.75	2.61	2.51	2.44	2.38	2.30	2.18	2.12	2.08	1.91
30	5.57	4.18	3.59	3.25	3.03	2.81	2.65	2.51	2.41	2.34	2.28	2.20	2.07	2.01	1.97	1.79
35	5.48	4.11	3.52	3.18	2.96	2.80	2.58	2.44	2.34	2.27	2.21	2.12	2.00	1.93	1.89	1.70
40	5.42	4.05	3.46	3.13	2.90	2.74	2.53	2.39	2.29	2.21	2.15	2.07	1.94	1.88	1.83	1.64
45	5.38	4.01	3.42	3.09	2.86	2.70	2.49	2.35	2.25	2.17	2.11	2.03	1.90	1.83	1.79	1.59
50	5.34	3.97	3.39	3.05	2.83	2.67	2.46	2.32	2.22	2.14	2.08	1.99	1.87	1.80	1.75	1.55
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.19	2.05	1.94	1.87	1.80	1.71	1.57	1.48	1.43	1.00

I valori $F_{0.05}, F_{0.025}, F_{0.005}$ si possono calcolare tramite la formula:

$$F_\alpha(\nu_1, \nu_2) = \frac{1}{F_{1-\alpha}(\nu_2, \nu_1)}$$

Distribuzione di FISHER.
Valori dei frattili $F_p(\nu_1, \nu_2)$
[ν_1, ν_2 gradi di libertà dei
campioni], dati da:

$$P = \int_0^{F_p} f(F) dF = 0.99$$



ν_2	1	2	3	4	5	6	8	10	12	14	16	20	30	40	50	∞
1	4052	5000	5403	5625	5764	5859	5982	6056	6106	6143	6169	6204	6261	6287	6303	6366
2	98.49	99.01	99.18	99.31	99.30	99.33	99.32	99.49	99.41	99.46	99.47	99.43	99.42	99.41	44.59	99.50
3	34.11	30.83	29.46	28.70	28.23	27.90	27.49	27.23	27.05	26.92	26.83	26.69	26.51	26.41	26.30	26.13
4	21.19	17.99	16.69	15.98	15.53	15.21	14.80	14.55	14.38	14.25	14.15	14.02	13.84	13.75	13.69	13.46
5	16.26	13.27	12.06	11.35	10.97	10.67	10.29	10.05	9.89	9.77	9.68	9.55	9.38	9.29	9.24	9.02
6	13.74	10.92	9.78	9.15	8.75	8.47	8.10	7.87	7.72	7.61	7.52	7.40	7.23	7.14	7.09	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.84	6.62	6.47	6.36	6.28	6.16	5.99	5.91	5.86	5.65
8	11.25	8.65	7.59	7.01	6.63	6.37	6.03	5.82	5.67	5.56	5.48	5.36	5.20	5.12	5.07	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.47	5.26	5.11	5.01	4.92	4.81	4.65	4.57	4.52	4.31
10	10.04	7.56	6.55	6.00	5.64	5.39	5.06	4.85	4.71	4.60	4.52	4.41	4.25	4.17	4.12	3.91
11	9.64	7.21	6.22	5.67	5.32	5.07	4.74	4.54	4.40	4.29	4.21	4.10	3.94	3.86	3.81	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.50	4.30	4.16	4.05	3.97	3.86	3.70	3.62	3.57	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.30	4.10	3.96	3.86	3.78	3.66	3.51	3.43	3.38	3.17
14	8.86	6.51	5.56	5.04	4.70	4.46	4.14	3.94	3.80	3.70	3.62	3.51	3.35	3.27	3.22	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.00	3.80	3.67	3.56	3.49	3.37	3.21	3.13	3.08	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	3.89	3.69	3.55	3.45	3.37	3.26	3.10	3.02	2.97	2.75
17	8.40	6.11	5.19	4.67	4.34	4.10	3.79	3.59	3.46	3.35	3.21	3.16	3.00	2.92	2.87	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.71	3.51	3.37	3.27	3.19	3.08	2.92	2.84	2.78	2.51
19	8.18	5.93	5.01	4.50	4.17	3.94	3.63	3.43	3.30	3.19	3.12	3.00	2.84	2.16	2.71	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.56	3.37	3.23	3.13	3.05	2.94	2.78	2.69	2.64	2.42
25	7.77	5.57	4.68	4.18	3.86	3.63	3.32	3.13	2.99	2.89	2.81	2.10	2.54	2.45	2.40	2.17
30	7.56	5.35	4.51	4.02	3.70	3.47	3.17	2.98	2.84	2.74	2.66	2.55	2.39	2.30	2.25	2.01
35	7.42	5.27	4.40	3.91	3.59	3.37	3.07	2.88	2.74	2.64	2.56	2.44	2.28	2.19	2.14	1.09
40	7.31	5.18	4.31	3.83	3.51	3.29	2.99	2.80	2.66	2.56	2.48	2.37	2.20	2.11	2.06	1.80
45	7.23	5.11	4.25	3.77	3.45	3.23	2.94	2.14	2.61	2.51	2.43	2.31	2.14	2.05	2.00	1.74
50	7.17	5.06	4.20	3.72	3.41	3.19	2.89	2.70	2.56	2.46	2.38	2.27	2.10	2.01	1.95	1.68
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.51	2.32	2.18	2.08	2.00	1.88	1.70	1.59	1.52	1.00

I valori $F_{0.05}, F_{0.025}, F_{0.005}$ si possono calcolare tramite la formula:

$$F_{\alpha}(\nu_1, \nu_2) = \frac{1}{F_{1-\alpha}(\nu_2, \nu_1)}$$